GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:35:52 ; Search time 8.64516 Seconds
(without alignments)
73.441 Million cell updates/sec

09-904-CAC-1

Title: US-~~10-050-611-2~~
Perfect score: 21
Sequence: 1 RGDA 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	4	13	AAR25315	Cell contact inhib
2	21	100.0	4	22	AAB86859	Transport molecule
3	21	100.0	4	23	AAE28393	Thrombo-spondin 1
4	21	100.0	4	23	AAE20157	Human thrombin pep
5	21	100.0	4	23	AAU78374	Thrombin peptide d
6	21	100.0	4	23	AAM50856	Thrombin receptor
7	21	100.0	5	13	AAR24517	Platelet antagonis
8	21	100.0	5	20	AAY17781	Human thrombospond
9	21	100.0	5	22	AAB72600	Thrombin-induced p
10	21	100.0	6	11	AAR04871	Peptide from fibro
11	21	100.0	6	12	AAR11506	Cell attachment pr
12	21	100.0	7	23	ABP48385	Zinc finger protei
13	21	100.0	7	23	ABP48594	Zinc finger protei
14	21	100.0	7	23	ABP48597	Zinc finger protei
15	21	100.0	7	23	ABP48600	Zinc finger protei
16	21	100.0	7	23	ABP48603	Zinc finger protei
17	21	100.0	7	23	ABP48606	Zinc finger protei
18	21	100.0	7	23	ABP48609	Zinc finger protei
19	21	100.0	7	23	ABP48672	Zinc finger protei
20	21	100.0	7	23	ABP48868	Zinc finger protei
21	21	100.0	7	23	ABP48883	Zinc finger protei
22	21	100.0	7	23	ABP49111	Zinc finger protei
23	21	100.0	7	23	ABP49114	Zinc finger protei
24	21	100.0	7	23	ABP49409	Zinc finger protei
25	21	100.0	7	23	ABP49436	Zinc finger protei
26	21	100.0	7	23	ABP49439	Zinc finger protei
27	21	100.0	7	23	ABP49445	Zinc finger protei
28	21	100.0	7	23	ABP49448	Zinc finger protei
29	21	100.0	7	23	ABP49547	Zinc finger protei
30	21	100.0	7	23	ABP49631	Zinc finger protei
31	21	100.0	7	23	ABP49634	Zinc finger protei
32	21	100.0	7	23	ABP49652	Zinc finger protei
33	21	100.0	7	23	ABP49655	Zinc finger protei
34	21	100.0	7	23	ABP49679	Zinc finger protei
35	21	100.0	7	23	ABP49685	Zinc finger protei
36	21	100.0	7	23	ABP50170	Zinc finger protei
37	21	100.0	7	23	ABP50194	Zinc finger protei
38	21	100.0	7	23	ABP50270	Zinc finger protei
39	21	100.0	7	23	ABP50310	Zinc finger protei
40	21	100.0	7	23	ABP50315	Zinc finger protei
41	21	100.0	7	23	ABP50411	Zinc finger protei
42	21	100.0	7	23	ABP50964	Zinc finger protei
43	21	100.0	7	23	ABP51010	Zinc finger protei
44	21	100.0	8	19	AAW48507	Integrin receptor
45	21	100.0	8	24	ABG72758	Human FNfn10 FG lo

ALIGNMENTS

```
RESULT 1
AAR25315
ID   AAR25315 standard; peptide; 4 AA.
XX
AC   AAR25315;
XX
DT   17-MAR-1993  (first entry)
XX
DE   Cell contact inhibitor generic peptide #4.
XX
KW   Cyclic peptide; cell contact inhibitor; hydrolytic enzyme.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   2
FT                   /label= MeGly
XX
PN   JP04264097-A.
XX
PD   18-SEP-1992.
XX
PF   16-FEB-1991;   91JP-0044386.
XX
PR   16-FEB-1991;   91JP-0044386.
XX
PA   (ASAG ) ASAHI GLASS CO LTD.
XX
DR   WPI; 1992-361922/44.
XX
PT   Peptide derivs. as contact inhibitor for animal cells - comprise
PT   synthesised cyclic peptide and have portion of aminoacid sequence
PT   of arginine-N-methyl:glycine-aspartic acid
XX
PS   Disclosure; Page 3; 6pp; Japanese.
XX
CC   The sequences given in AAR25311-19 are cyclic peptides which act as
CC   contact inhibitors of animal cells.  They are resistant to
CC   decomposition by hydrolytic enzymes and can be maintained at high
CC   levels of activity for a long period in vivo.  The peptides are
CC   cyclic and may have 1-16 pref. 1-4 amino acids.
XX
SQ   Sequence   4 AA;

  Query Match            100.0%;  Score 21;  DB 13;  Length 4;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          1 RGDA 4


RESULT 2
AAB86859
ID   AAB86859 standard; peptide; 4 AA.
XX
AC   AAB86859;
XX
DT   28-NOV-2001  (first entry)
XX
DE   Transport molecule/ligand binding-associated peptide #5.
XX
KW   Transport molecule; ligand; cancer treatment; autoimmune disease;
KW   inflammation; infection.
XX
OS   Synthetic.
XX
PN   WO200168142-A1.
XX
PD   20-SEP-2001.
XX
PF   13-MAR-2001; 2001WO-EP02833.
XX
PR   13-MAR-2000; 2000DE-1012120.
XX
PA   (KTBT-) KTB TUMORFORSCHUNGS GMBH.
XX
PI   Kratz F;
XX
DR   WPI; 2001-589998/66.
XX
PT   New ligand, comprising therapeutic or diagnostic agent bonded
PT   non-covalently with substance having high affinity to transport
PT   molecule  -
XX
PS   Disclosure; Page 39; 74pp; German.
XX
CC   This invention describes novel ligands which bind to transport molecules,
CC   comprising a therapeutic and/or diagnostic agent (A) non-covalently
CC   bonded via a linkage cleavable in vivo depending on pH and/or
CC   enzymatically with a substance (B) having an association constant KA to a
CC   transport molecule of above 10^3 M^-^1, is new. The medicaments are
CC   especially useful for the treatment of cancers, autoimmune diseases,
CC   acute and chronic inflammation and infections caused by viruses or
CC   microorganisms. The diagnostic kits are useful for the detection of these
CC   illnesses and for the detection of the transport molecule and/or its
CC   distribution in vivo. The ligands have excellent solubility in the medium
CC   at the site of action and are easy and inexpensive to covert into
CC   adducts, as the interaction with the transport material is physical.
CC   AAB86843-AAB86920 represent peptides used to illustrate the
CC   method of the invention.
XX
SQ   Sequence   4 AA;

  Query Match            100.0%;  Score 21;  DB 22;  Length 4;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          1 RGDA 4
```

RESULT 3
AAE28393
ID AAE28393 standard; peptide; 4 AA.
XX
AC AAE28393;
XX
DT 27-DEC-2002 (first entry)
XX
DE Thrombo-spondin 1 RGD cell binding region.
XX
KW Tat region; nucleic acid-binding group; cell transfection system;
KW gene therapy; cancer; thrombo-spondin 1.
XX
OS Unidentified.
XX
PN US6376248-B1.
XX
PD 23-APR-2002.
XX
PF 16-MAR-1998; 98US-0039780.
XX
PR 14-MAR-1997; 97US-0818200.
XX
PA (LIFE-) LIFE TECHNOLOGIES INC.
XX
PI Hawley-Nelson P, Lan J, Shih P, Jessee JA, Schifferli KP;
PI Gebeyehu G, Ciccarone VC, Evans KL;
XX
DR WPI; 2002-680647/73.
XX
PT New peptide comprising Tat sequence linked to nucleic acid-binding
PT group, useful, e.g. in gene therapy, for improving cell-transfection
PT efficiency -
XX
PS Example 1; Column 65; 108pp; English.
XX
CC The invention relates to a peptide comprising Tat sequence linked to
CC nucleic acid-binding group. Peptides of the invention are used as
CC components of a cell transfection system particularly for gene therapy
CC (especially of cancer). The present sequence is thrombo-spondin 1 RGD
CC cell binding region. This peptide is used in the exemplification of
CC the invention.
XX
SQ Sequence 4 AA;

  Query Match   100.0%; Score 21; DB 23; Length 4;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   1 RGDA 4

RESULT 4
AAE20157
ID AAE20157 standard; peptide; 4 AA.
XX
AC AAE20157;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human thrombin peptide.
XX
KW Cartilage growth; cartilage repair; arthritic joint; traumatic injury;
KW non-proteolytically activated thrombin receptor; NPAR; chondrocyte;
KW therapy; implantation; thrombin peptide; human.
XX
OS Homo sapiens.
XX
PN WO200207748-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US22668.
XX
PR 20-JUL-2000; 2000US-219800P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH, Crowther RS, Stiernberg J, Bergmann J;
XX
DR WPI; 2002-268953/31.
XX
PT Stimulating growth and repair of cartilage, useful for treating e.g.
PT arthritis, by local administration of an agonist of non-proteolytically
PT activated thrombin receptor -
XX
PS Claim 10; Page 25; 28pp; English.
XX
CC The invention relates to a method of stimulating growth and repair of
CC cartilage. The method involves administering to the site, an agonist
CC of non-proteolytically activated thrombin receptor (NPAR). The method
CC is used in human or veterinary medicine for the treatment of arthritic
CC joints and damage/loss of cartilage caused by traumatic injury. Also
CC chondrocytes may be cultured in presence of NPAR agonist to provide
CC cells for implantation at sites requiring growth/repair of cartilage.
CC The present sequence is human thrombin peptide. The derivatives of
CC thrombin peptide which serves as a NPAR agonist.
XX
SQ Sequence 4 AA;

  Query Match   100.0%; Score 21; DB 23; Length 4;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   1 RGDA 4

RESULT 5
AAU78374
ID AAU78374 standard; Peptide; 4 AA.

XX
AC AAU78374;
XX
DT 18-JUN-2002 (first entry)
XX
DE Thrombin peptide derivative #1.
XX
KW Thrombin; osteopathic; receptor; agonist; bone growth stimulation;
KW osteoinduction; farm animal; companion animal; laboratory animal;
KW bone graft; segmental bone gap; bone void; non-union fracture.
XX
OS Synthetic.
XX
PN WO200205836-A2.
XX
PD 24-JAN-2002.
XX
PF 18-JUL-2001; 2001WO-US22641.
XX
PR 19-JUL-2000; 2000US-219300P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH, Crowther RS, Simmons DJ, Yang J, Redin WR;
XX
DR WPI; 2002-303796/34.
XX
PT Stimulating bone growth at a site in a subject in need of
PT osteoinduction, such as a site of bone graft, segmental bone gap, bone
PT void or non-union structure, by administering agonist of activated
PT thrombin receptor -
XX
PS Claim 9; Page 22; 27pp; English.
XX
CC The invention describes a method of stimulating bone growth at a site
CC in a subject in need of osteoinduction. The method involves administering
CC an agonist to stimulate bone growth at a site in a subject (e.g. a farm
CC animal, companion animal or laboratory animal), in need of
CC osteoinduction, such as the site in need of a bone graft in a subject, a
CC segmental bone gap, a bone void or a non-union fracture. This sequence
CC represents a thrombin peptide derivative obtained from a serine
CC esterase that can stimulate or activate the non-proteolytically
CC activated thrombin receptor.
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 23; Length 4;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
||||
Db 1 RGDA 4

RESULT 6
AAM50856

ID AAM50856 standard; Peptide; 4 AA.
XX
AC AAM50856;
XX
DT 01-MAY-2002 (first entry)
XX
DE Thrombin receptor binding domain used for cardiac tissue repair.
XX
KW Thrombin receptor binding domain; thrombin; revascularisation;
KW vascular occlusion; tissue repair; vulnerary; vasotropic; cardiant;
KW angiogenesis; restenosis; therapy; human.
XX
OS Homo sapiens.
XX
PN WO200204008-A2.
XX
PD 17-JAN-2002.
XX
PF 12-JUL-2001; 2001WO-US21944.
XX
PR 12-JUL-2000; 2000US-217583P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH;
XX
DR WPI; 2002-179665/23.
XX
PT Promoting cardiac tissue repair, stimulating revascularisation,
PT stimulating vascular endothelial cell proliferation, and inhibiting
PT vascular occlusion by using angiogenic thrombin derivative peptide -
XX
PS Claim 2; Page 19; 24pp; English.
XX
CC The present sequence is that of a thrombin receptor binding domain
CC peptide that is used in a claimed method for promoting cardiac
CC tissue repair. The method involves administering an angiogenic
CC thrombin-derived peptide. The peptide comprises the present
CC thrombin receptor binding domain together with a serine esterase
CC conserved sequence (see AAM50857), or preferably a peptide (see
CC AAM50858) which includes both these sequences. The thrombin-derived
CC peptide is administered during or following cardiac surgery by
CC injection into cardiac tissue, and may be formulated as a sustained
CC release formulation. It is used in claimed methods of stimulating
CC revascularisation, stimulating vascular endothelial cell
CC proliferation, inhibiting vascular occlusion, and inhibiting
CC restenosis following balloon angioplasty, in which case the
CC peptide may be coated onto the catheter.
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 23; Length 4;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
||||

```
Db          1 RGDA 4


RESULT 7
AAR24517
ID   AAR24517 standard; Protein; 5 AA.
XX
AC   AAR24517;
XX
DT   02-DEC-1992  (first entry)
XX
DE   Platelet antagonist pepetide 4.
XX
KW   Clinical effect; antagonist.
XX
OS   Synthetic.
XX
PN   JP04134096-A.
XX
PD   07-MAY-1992.
XX
PF   21-SEP-1990;   90JP-0253849.
XX
PR   21-SEP-1990;   90JP-0253849.
XX
PA   (SEGK ) SEIKAGAKU KOGYO CO LTD.
XX
DR   WPI; 1992-204525/25.
XX
PT   New peptide(s) comprising arginine-glycine-asparagine and
PT   hyaluronic acid - useful as platelet antagonists with higher
PT   activity than arginine-glycine-asparagine-valine
XX
PS   Disclosure; Page 5; 10pp; Japanese.
XX
CC   The sequences given in AAR24514-8 are peptides which are useful as
CC   platelet antagonists.  These peptides have higher activity than the
CC   conventional peptide of Arg-Gly-Asp-Val.  These peptides have a
CC   clinical effect at a lower dose, dosage is 2.5-5.0 mg/kg/day.
XX
SQ   Sequence   5 AA;

  Query Match             100.0%;  Score 21;  DB 13;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          2 RGDA 5


RESULT 8
AAY17781
ID   AAY17781 standard; peptide; 5 AA.
XX
AC   AAY17781;
XX
DT   12-AUG-1999  (first entry)
XX
DE   Human thrombospondin-1 type III repeat peptide.
XX
KW   Human; thrombospondin; HIV; infection; inhibition; chemokine;
KW   contraceptive.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9926649-A1.
XX
PD   03-JUN-1999.
XX
PF   24-NOV-1998;   98WO-US24905.
XX
PR   20-MAR-1998;   98US-0078873.
PR   25-NOV-1997;   97US-0066294.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Crombie AR,  Laurence JC,  Nachman RL;
XX
DR   WPI; 1999-370856/31.
XX
PT   Suppressing infectivity of human immune deficiency virus
XX
PS   Example 2; Page 33; 67pp; English.
XX
CC   The present invention describes a method for suppressing infectivity of
CC   human immunodeficiency virus (HIV) by treating the virus, or its target
CC   cell, with a thrombospondin or thrombospondin analogue. Thrombospondin
CC   blocks binding of HIV to its cellular receptors. Thrombospondin or its
CC   analogues can used to prevent infection by HIV, in both contraceptive
CC   and non-contraceptive compositions/devices. They are already known to
CC   reduce infectivity of some bacteria and protozoa. The present sequence
CC   represents a human thrombospondin-1 type III repeat peptide.
XX
SQ   Sequence   5 AA;

  Query Match             100.0%;  Score 21;  DB 20;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          2 RGDA 5


RESULT 9
AAB72600
ID   AAB72600 standard; Peptide; 5 AA.
XX
AC   AAB72600;
XX
DT   09-MAY-2001  (first entry)
XX
```

DE Thrombin-induced platelet activator antagonist #39.
XX
KW Platelet aggregation inhibitor; thrombin activation inhibitor;
KW protease activated receptor 1; PAR1; platelet activation inhibitor;
KW thrombosis; acute coronary syndrome.
XX
OS Unidentified.
XX
PN WO200112656-A1.
XX
PD 22-FEB-2001.
XX
PF 17-AUG-2000; 2000WO-US40669.
XX
PR 17-AUG-1999; 99US-0375808.
XX
PA (THRO-) THROMGEN INC.
XX
PI Schmaier AH, Hasan AAK;
XX
DR WPI; 2001-226546/23.
XX
PT Inhibiting thrombin activation in human cell expressing protease
PT activated receptor 1 (PAR1), comprises contacting mixtures of thrombin
PT and human cell expressing PAR1, with a peptide that inhibits platelet
PT activation -
XX
PS Claim 8; Page 26; 49pp; English.
XX
CC The present invention relates to a method for inhibiting thrombin
CC activation in a human cell expressing protease activated receptor 1
CC (PAR1). The method involves using peptides (e.g. the present peptide)
CC that inhibit platelet activation. The method is useful for preventing
CC thrombosis and platelet aggregation. The method can be used for patients
CC with acute coronary syndromes (e.g. cresendo angina, myocardial
CC infarction) and for individuals who have acute coronary syndromes and
CC receive percutaneous transluminal coronary angioplasty with an artificle
CC stent placement.
XX
SQ Sequence 5 AA;

  Query Match   100.0%; Score 21; DB 22; Length 5;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   1 RGDA 4

RESULT 10
AAR04871
ID AAR04871 standard; peptide; 6 AA.
XX
AC AAR04871;
XX
DT 25-MAR-2003 (updated)

DT 25-SEP-1989 (first entry)
XX
DE Peptide from fibronectin.
XX
KW Fibronectin; cell attachment; cell detachment; fermentation; therapy.
XX
OS synthetic.
XX
PN US4879237-A.
XX
PD 07-NOV-1989.
XX
PF 24-MAY-1985; 85US-0738078.
XX
PR 24-MAY-1985; 85US-0738078.
XX
PA (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI Ruoslahti EI, Hayman EG, Pierschbacher MD;
XX
DR WPI; 1990-154405/20.
XX
PT Synthetic peptide(s) from fibronectin- used in control of cell attachment
PT and detachment
XX
PS Claim 1; page 10; 13pp; English.
XX
CC This polypeptide mediates the attachment of animal cells to substrates.
CC The substrate (I) is contacted with cells and with a soln. contg. this
CC polypeptide. This attachment can be prevented in addition to detaching
CC the cells from (I) once attached. Applications are in eg fermentation,
CC cell line prepn., diagnosis and therapy.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 6 AA;

  Query Match   100.0%; Score 21; DB 11; Length 6;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   2 RGDA 5

RESULT 11
AAR11506
ID AAR11506 standard; Protein; 6 AA.
XX
AC AAR11506;
XX
DT 12-JUN-1991 (first entry)
XX
DE Cell attachment promoting peptide.
XX
KW Fibrin; aggregation.

XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Active-site 2..4
XX
PN US4988621-A.
XX
PD 29-JAN-1991.
XX
PF 10-DEC-1987; 87US-0131130.
XX
PR 10-DEC-1987; 87US-0131130.
PR 24-MAY-1985; 85US-0738078.
XX
PA (JOLL-) LA JOLLA CANCER FOU.
XX
PI Ruoslahti EI, Hayman EG, Pierschbacher MD;
XX
DR WPI; 1991-116404/16.
XX
PT Peptide(s) contg. arginine-glycine-aspartic acid sequence - used
PT to prevent and reverse cell attachment or to promote cell
PT aggregation.
XX
PS Disclosure; Page 8; 12pp; English.
XX
CC The peptide, or shorter versions contg. the RGD active site from
CC fibronectin, can be used to prevent and reverse attachment of cells
CC to substrates. This can be used in cell prodn., fermentation, cell
CC line prepn., cell matrix prodn., diagnostics and therapy. The
CC peptide can be used for eg mobilisation of bone marrow cells;
CC prevention and reversal of attachment of disseminated tumour cells
CC locally such as in the case of an operation performed in the peri-
CC toneal cavity, to prevent adhesions and scar formations locally as
CC in the case of eye operations, for prophylactic inhibiton of E. coli
CC binding to epithelial cells of the urinary tract or intestine,
CC diagnosis and treatment of E. coli related infections, and
CC identification of various pathogenic bacterial strains. The
CC peptide is pref. prepd. by solid phase synthesis.
CC See also AAR11505
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 21; DB 12; Length 6;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
||||
Db 2 RGDA 5

RESULT 12
ABP48385
ID ABP48385 standard; Peptide; 7 AA.
XX
AC ABP48385;
XX
DT 28-AUG-2002 (first entry)
XX
DE Zinc finger protein related peptide motif SEQ ID NO:289.
XX
KW Zinc finger protein; ZFP; DNA binding protein; zinc finger.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200242459-A2.
XX
PD 30-MAY-2002.
XX
PF 20-NOV-2001; 2001WO-US43438.
XX
PR 20-NOV-2000; 2000US-0716637.
XX
PA (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI Liu Q;
XX
DR WPI; 2002-500284/53.
XX
PT New zinc finger protein that binds to target site, useful in studying
PT gene function and for human therapeutics and plant engineering,
PT comprises first, second and third zinc fingers, ordered from N- to
PT C-terminus -
XX
PS Example 1; Page 37; 81pp; English.
XX
CC The present invention describes a zinc finger protein (I) that binds to
CC a target site, comprising a first (F1), a second (F2), and a third (F3)
CC zinc finger, ordered F1, F2, F3 from N-terminus to C-terminus, where the
CC target site comprises, in 3'-5' direction, a first (S1), a second (S2),
CC and a third (S3) target subsite. Also described are: (1) a polypeptide
CC (II) comprising (I); (2) a polynucleotide (III) encoding (I) or (II); and
CC (3) designing (M) (I) involves selecting the F1 zinc finger such that
CC it binds to the S1 target subsite, selecting the F2 zinc finger such
CC that it binds to the S2 target subsite, and selecting the F3 zinc
CC finger such that it binds to the S3 target subsite, thus designing (I)
CC that binds to a target site. (I) is useful for recognition of triplet
CC target subsites having the nucleotide G in the 5'-most position of the
CC subsite. (I) is useful in studying gene function, and for human
CC therapeutics and plant engineering. (I), (II) or (III) is useful in
CC therapeutic methods to modulate the expression of a target region within
CC a subject, in diagnostic methods for sequence specific detection of
CC target nucleic acid in a sample, and in assays to determined the
CC phenotype and function of gene expression. (I) has improved affinity
CC and specificity for their target sequences, as well as enhanced
CC biological activity. ABQ71213 to ABQ72214 and ABP48191 to ABP51230
CC represent DNA target sequences and zinc finger peptides which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 7 AA;

```
  Query Match           100.0%;  Score 21;  DB 23;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          1 RGDA 4


RESULT 13
ABP48594
ID   ABP48594 standard; Peptide; 7 AA.
XX
AC   ABP48594;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Zinc finger protein related peptide motif SEQ ID NO:670.
XX
KW   Zinc finger protein; ZFP; DNA binding protein; zinc finger.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200242459-A2.
XX
PD   30-MAY-2002.
XX
PF   20-NOV-2001; 2001WO-US43438.
XX
PR   20-NOV-2000; 2000US-0716637.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Liu Q;
XX
DR   WPI; 2002-500284/53.
XX
PT   New zinc finger protein that binds to target site, useful in studying
PT   gene function and for human therapeutics and plant engineering,
PT   comprises first, second and third zinc fingers, ordered from N- to
PT   C-terminus  -
XX
PS   Example 1; Page 40; 81pp; English.
XX
CC   The present invention describes a zinc finger protein (I) that binds to
CC   a target site, comprising a first (F1), a second (F2), and a third (F3)
CC   zinc finger, ordered F1, F2, F3 from N-terminus to C-terminus, where the
CC   target site comprises, in 3'-5' direction, a first (S1), a second (S2),
CC   and a third (S3) target subsite. Also described are: (1) a polypeptide
CC   (II) comprising (I); (2) a polynucleotide (III) encoding (I) or (II); and
CC   (3) designing (M) (I) involves selecting the F1 zinc finger such that
CC   it binds to the S1 target subsite, selecting the F2 zinc finger such
CC   that it binds to the S2 target subsite, and selecting the F3 zinc
CC   finger such that it binds to the S3 target subsite, thus designing (I)
CC   that binds to a target site. (I) is useful for recognition of triplet
CC   target subsites having the nucleotide G in the 5'-most position of the
CC   subsite. (I) is useful in studying gene function, and for human
CC   therapeutics and plant engineering. (I), (II) or (III) is useful in
CC   therapeutic methods to modulate the expression of a target region within
CC   a subject, in diagnostic methods for sequence specific detection of
CC   target nucleic acid in a sample, and in assays to determined the
CC   phenotype and function of gene expression. (I) has improved affinity
CC   and specificity for their target sequences, as well as enhanced
CC   biological activity. ABQ71213 to ABQ72214 and ABP48191 to ABP51230
CC   represent DNA target sequences and zinc finger peptides which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence   7 AA;

  Query Match           100.0%;  Score 21;  DB 23;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          1 RGDA 4


RESULT 14
ABP48597
ID   ABP48597 standard; Peptide; 7 AA.
XX
AC   ABP48597;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Zinc finger protein related peptide motif SEQ ID NO:671.
XX
KW   Zinc finger protein; ZFP; DNA binding protein; zinc finger.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200242459-A2.
XX
PD   30-MAY-2002.
XX
PF   20-NOV-2001; 2001WO-US43438.
XX
PR   20-NOV-2000; 2000US-0716637.
XX
PA   (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI   Liu Q;
XX
DR   WPI; 2002-500284/53.
XX
PT   New zinc finger protein that binds to target site, useful in studying
PT   gene function and for human therapeutics and plant engineering,
PT   comprises first, second and third zinc fingers, ordered from N- to
PT   C-terminus  -
XX
PS   Example 1; Page 40; 81pp; English.
```

XX
CC The present invention describes a zinc finger protein (I) that binds to
CC a target site, comprising a first (F1), a second (F2), and a third (F3)
CC zinc finger, ordered F1, F2, F3 from N-terminus to C-terminus, where the
CC target site comprises, in 3'-5' direction, a first (S1), a second (S2),
CC and a third (S3) target subsite. Also described are: (1) a polypeptide
CC (II) comprising (I); (2) a polynucleotide (III) encoding (I) or (II); and
CC (3) designing (M) (I) involves selecting the F1 zinc finger such that
CC it binds to the S1 target subsite, selecting the F2 zinc finger such
CC that it binds to the S2 target subsite, and selecting the F3 zinc
CC finger such that it binds to the S3 target subsite, thus designing (I)
CC that binds to a target site. (I) is useful for recognition of triplet
CC target subsites having the nucleotide G in the 5'-most position of the
CC subsite. (I) is useful in studying gene function, and for human
CC therapeutics and plant engineering. (I), (II) or (III) is useful in
CC therapeutic methods to modulate the expression of a target region within
CC a subject, in diagnostic methods for sequence specific detection of
CC target nucleic acid in a sample, and in assays to determined the
CC phenotype and function of gene expression. (I) has improved affinity
CC and specificity for their target sequences, as well as enhanced
CC biological activity. ABQ71213 to ABQ72214 and ABP48191 to ABP51230
CC represent DNA target sequences and zinc finger peptides which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 21; DB 23; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
||||
Db 1 RGDA 4

RESULT 15
ABP48600
ID ABP48600 standard; Peptide; 7 AA.
XX
AC ABP48600;
XX
DT 28-AUG-2002 (first entry)
XX
DE Zinc finger protein related peptide motif SEQ ID NO:672.
XX
KW Zinc finger protein; ZFP; DNA binding protein; zinc finger.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200242459-A2.
XX
PD 30-MAY-2002.
XX
PF 20-NOV-2001; 2001WO-US43438.
XX
PR 20-NOV-2000; 2000US-0716637.
XX
PA (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI Liu Q;
XX
DR WPI; 2002-500284/53.
XX
PT New zinc finger protein that binds to target site, useful in studying
PT gene function and for human therapeutics and plant engineering,
PT comprises first, second and third zinc fingers, ordered from N- to
PT C-terminus -
XX
PS Example 1; Page 40; 81pp; English.
XX
CC The present invention describes a zinc finger protein (I) that binds to
CC a target site, comprising a first (F1), a second (F2), and a third (F3)
CC zinc finger, ordered F1, F2, F3 from N-terminus to C-terminus, where the
CC target site comprises, in 3'-5' direction, a first (S1), a second (S2),
CC and a third (S3) target subsite. Also described are: (1) a polypeptide
CC (II) comprising (I); (2) a polynucleotide (III) encoding (I) or (II); and
CC (3) designing (M) (I) involves selecting the F1 zinc finger such that
CC it binds to the S1 target subsite, selecting the F2 zinc finger such
CC that it binds to the S2 target subsite, and selecting the F3 zinc
CC finger such that it binds to the S3 target subsite, thus designing (I)
CC that binds to a target site. (I) is useful for recognition of triplet
CC target subsites having the nucleotide G in the 5'-most position of the
CC subsite. (I) is useful in studying gene function, and for human
CC therapeutics and plant engineering. (I), (II) or (III) is useful in
CC therapeutic methods to modulate the expression of a target region within
CC a subject, in diagnostic methods for sequence specific detection of
CC target nucleic acid in a sample, and in assays to determined the
CC phenotype and function of gene expression. (I) has improved affinity
CC and specificity for their target sequences, as well as enhanced
CC biological activity. ABQ71213 to ABQ72214 and ABP48191 to ABP51230
CC represent DNA target sequences and zinc finger peptides which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 21; DB 23; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
||||
Db 1 RGDA 4

Search completed: February 11, 2004, 14:53:24
Job time : 10.6452 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:49:07 ; Search time 2.70968 Seconds
(without alignments)
141.963 Million cell updates/sec

Title: US-10-050-611-1
Perfect score: 21
Sequence: 1 RGDA 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	19	2	A34467	36K microfibril-as
2	21	100.0	39	2	A36453	decorsin - leech (
3	21	100.0	45	2	G82812	hypothetical prote
4	21	100.0	49	2	S70093	hypothetical prote
5	21	100.0	52	2	S19623	ornatin C - leech
6	21	100.0	57	2	E70535	hypothetical prote
7	21	100.0	68	2	AG3217	hypothetical prote
8	21	100.0	74	2	S62570	60s ribosomal prot
9	21	100.0	76	2	I39905	trp RNA-binding pr
10	21	100.0	79	2	B90870	hypothetical prote
11	21	100.0	79	2	G85748	unknown protein en
12	21	100.0	79	2	E64884	ydaQ protein - Esc
13	21	100.0	80	2	S68677	cytochrome c551 -
14	21	100.0	88	2	H82662	conserved hypothet
15	21	100.0	89	2	I68553	cell surface glyco
16	21	100.0	90	2	E82562	hypothetical prote
17	21	100.0	93	2	AH0620	probable prophage
18	21	100.0	95	2	E82696	hypothetical prote
19	21	100.0	96	2	G84240	hypothetical prote
20	21	100.0	96	2	D83771	hypothetical prote
21	21	100.0	97	2	A71054	ribosomal protein
22	21	100.0	97	2	C75089	ribosomal protein
23	21	100.0	97	2	E82962	hypothetical prote
24	21	100.0	98	2	S01566	hypothetical prote
25	21	100.0	100	2	T30673	hypothetical prote
26	21	100.0	102	2	E75273	conserved hypothet
27	21	100.0	103	2	F70976	hypothetical prote
28	21	100.0	104	2	B72538	probable acylphosp
29	21	100.0	107	2	F90230	partial transposas
30	21	100.0	108	2	T51207	hypothetical prote
31	21	100.0	110	2	AC2787	conserved hypothet
32	21	100.0	110	2	E97566	hypothetical prote
33	21	100.0	115	2	S14024	hypothetical prote
34	21	100.0	115	2	C82479	hypothetical prote
35	21	100.0	116	2	D71832	ribosomal protein
36	21	100.0	116	2	D64681	ribosomal protein
37	21	100.0	117	2	B81255	50S ribosomal prot
38	21	100.0	121	2	I35719	phnQ protein - Esc
39	21	100.0	123	2	H75059	hypothetical prote
40	21	100.0	124	2	D84319	30S ribosomal prot
41	21	100.0	124	2	S62816	ribosomal protein
42	21	100.0	124	2	T03574	hypothetical prote
43	21	100.0	126	2	C86883	50S ribosomal prot
44	21	100.0	126	2	B72621	hypothetical prote
45	21	100.0	126	2	T37063	hypothetical prote

ALIGNMENTS

RESULT 1
A34467
36K microfibril-associated protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 18-Jun-1993
C;Accession: A34467
R;Kobayashi, R.; Tashima, Y.; Masuda, H.; Shozawa, T.; Numata, Y.; Miyauchi, K.; Hayakawa, T.
J. Biol. Chem. 264, 17437-17444, 1989
A;Title: Isolation and characterization of a new 36-kDa microfibril-associated glycoprotein from porcine aorta.
A;Reference number: A34467; MUID:90008913; PMID:2793866
A;Accession: A34467
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-19 <KOB>

Query Match 100.0%; Score 21; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 60;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RGDA 4
              ||||
Db          5 RGDA 8
```

RESULT 2
A36453
decorsin - leech (Macrobdella decora)
C;Species: Macrobdella decora
C;Date: 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 30-Sep-1993
C;Accession: A36453
R;Seymour, J.L.; Henzel, W.J.; Nevins, B.; Stults, J.T.; Lazarus, R.A.
J. Biol. Chem. 265, 10143-10147, 1990
A;Title: Decorsin. A potent glycoprotein IIb-IIIa antagonist and platelet aggregation inhibitor from the leech Macrobdella decora.
A;Reference number: A36453; MUID:90277628; PMID:2351655
A;Accession: A36453
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-39 <SEY>

Query Match 100.0%; Score 21; DB 2; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RGDA 4
              ||||
Db         31 RGDA 34
```

RESULT 3
G82812
hypothetical protein XF0386 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82812
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: G82812
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-45 <SIM>
A;Cross-references: GB:AE003890; GB:AE003849; NID:g9105215; PIDN:AAF83196.1; GSPDB:GN00128; XFSC:XF0386
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0386

Query Match 100.0%; Score 21; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RGDA 4
              ||||
Db         19 RGDA 22
```

RESULT 4
S70093
hypothetical protein (orf49) - Amycolatopsis methanolica
C;Species: Amycolatopsis methanolica
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
C;Accession: S70093
R;Vrijbloed, J.W.; Jelinkova, M.; Hessels, G.I.; Dijkhuizen, L.
Mol. Microbiol. 18, 21-31, 1995
A;Title: Identification of the minimal replicon of plasmid pMEA300 of the methylotrophic actinomycete Amycolatopsis methanolica.
A;Reference number: S70087; MUID:96154938; PMID:8596458
A;Accession: S70093
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-49 <VRI>
A;Cross-references: EMBL:L36679
C;Genetics:
A;Start codon: GTG

Query Match 100.0%; Score 21; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;

```
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         23 RGDA 26
```

```
RESULT 5
S19623
ornatin C - leech (Placobdella ornata)
C;Species: Placobdella ornata
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C;Accession: S19623
R;Mazur, P.; Henzel, W.J.; Seymour, J.L.; Lazarus, R.A.
Eur. J. Biochem. 202, 1073-1082, 1991
A;Title: Ornatins: potent glycoprotein IIb-IIIa antagonists and platelet
aggregation inhibitors from the leech Placobdella ornata.
A;Reference number: S19566; MUID:92111479; PMID:1765068
A;Accession: S19623
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-52 <MAZ>

  Query Match             100.0%;  Score 21;  DB 2;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         42 RGDA 45
```

```
RESULT 6
E70535
hypothetical protein Rv0666 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: E70535
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaia, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.;
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete
genome sequence.
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70535
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-57 <COL>
A;Cross-references: GB:Z95972; GB:AL123456; NID:g3261790; PIDN:CAB09391.1;
PID:e319190; PID:g2143295
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0666

  Query Match             100.0%;  Score 21;  DB 2;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         24 RGDA 27
```

```
RESULT 7
AG3217
hypothetical protein Atu5470 [imported] - Agrobacterium tumefaciens (strain C58,
Dupont) plasmid AT
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AG3217
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen,
Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee
Sr., D.; Chapman, P.; Clendenning, J.; Deatherage, G.; Gillet, W.; Grant, C.;
Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClelland, E.; Palmieri, A.;
Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.;
Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.;
Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey,
S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens
C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AG3217
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-68 <KUR>
A;Cross-references: GB:AE008687; PIDN:AAL46157.1; PID:g17743927; GSPDB:GN00188
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu5470
A;Genome: plasmid

  Query Match             100.0%;  Score 21;  DB 2;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         36 RGDA 39
```

```
RESULT 8
S62570
60s ribosomal protein l38 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: protein SPAC30D11.1
C;Species: Schizosaccharomyces pombe
C;Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 11-Jan-2000
```

C;Accession: S62570; T38587
R;Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A;Reference number: S62559
A;Accession: S62570
A;Molecule type: DNA
A;Residues: 1-74 <PEA>
A;Cross-references: EMBL:Z67961; NID:g1065887; PIDN:CAA91898.1; PID:g1065899
R;Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z21801
A;Accession: T38587
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-74 <PE2>
A;Cross-references: EMBL:Z67961; PIDN:CAA91898.1; GSPDB:GN00066;
SPDB:SPAC30D11.12
A;Experimental source: strain 972h-; cosmid c30D11
C;Genetics:
A;Gene: rpl38-2; SPAC30D11.12
A;Map position: 1L
A;Introns: 1/3; 64/1
C;Superfamily: rat ribosomal protein L38
C;Keywords: cytosol; protein biosynthesis; ribosome

  Query Match   100.0%; Score 21; DB 2; Length 74;
  Best Local Similarity 100.0%; Pred. No. 2.2e+02;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   17 RGDA 20

RESULT 9
I39905
trp RNA-binding protein - Bacillus pumilus
C;Species: Bacillus pumilus
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999
C;Accession: I39905
R;Hoffman, R.J.; Gollnick, P.
J. Bacteriol. 177, 839-842, 1995
A;Title: The mtrB gene of Bacillus pumilus encodes a protein with sequence and
functional homology to the trp RNA-binding attenuation protein (TRAP) of
Bacillus subtilis.
A;Reference number: I39904; MUID:95138053; PMID:7836324
A;Accession: I39905
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-76 <RES>
A;Cross-references: GB:L37879; NID:g598076; PIDN:AAA67544.1; PID:g598078
C;Genetics:
A;Gene: mtrB

  Query Match   100.0%; Score 21; DB 2; Length 76;
  Best Local Similarity 100.0%; Pred. No. 2.3e+02;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   58 RGDA 61

RESULT 10
B90870
hypothetical protein ECs1930 [imported] - Escherichia coli (strain O157:H7,
substrain RIMD 0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B90870
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.;
Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida,
T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara,
S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7
and genomic comparison with a laboratory strain K-12.
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90870
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-79 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35353.1; PID:g13361395; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1930

  Query Match   100.0%; Score 21; DB 2; Length 79;
  Best Local Similarity 100.0%; Pred. No. 2.4e+02;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   5 RGDA 8

RESULT 11
G85748
unknown protein encoded within prophage CP-933R [imported] - Escherichia coli
(strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: G85748
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose,
D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.;
Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis,
N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.;
Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85748
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-79 <STO>
A;Cross-references: GB:AE005174; NID:g12515406; PIDN:AAG56451.1; GSPDB:GN00145; UWGP:Z2414
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ydaQ

  Query Match 100.0%; Score 21; DB 2; Length 79;
  Best Local Similarity 100.0%; Pred. No. 2.4e+02;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
      ||||
Db 5 RGDA 8

RESULT 12
E64884
ydaQ protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: E64884
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: E64884
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-79 <BLAT>
A;Cross-references: GB:AE000232; GB:U00096; NID:g1787600; PIDN:AAC74428.1; PID:g1787608; UWGP:b1346
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ydaQ

  Query Match 100.0%; Score 21; DB 2; Length 79;
  Best Local Similarity 100.0%; Pred. No. 2.4e+02;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
      ||||
Db 5 RGDA 8

RESULT 13
S68677
cytochrome c551 - Chromatium vinosum
C;Species: Chromatium vinosum
C;Date: 25-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 04-Mar-2000
C;Accession: S68677
R;Samyn, B.; de Smet, L.; van Driessche, G.; Meyer, T.E.; Bartsch, R.G.; Cusanovich, M.A.; van Beeumen, J.J.
Eur. J. Biochem. 236, 689-696, 1996
A;Title: A high-potential soluble cytochrome c-551 from the purple phototrophic bacterium Chromatium vinosum is homologous to cytochrome c(8) from denitrifying pseudomonas.
A;Reference number: S68677; MUID:96195682; PMID:8612646
A;Accession: S68677
A;Molecule type: protein
A;Residues: 1-80 <SAM>
A;Experimental source: strain D
C;Superfamily: cytochrome c6; cytochrome c6 homology
C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidative phosphorylation
F;1-77/Domain: cytochrome c6 homology <CYC>
F;10,13/Binding site: heme (Cys) (covalent) #status predicted
F;14,59/Binding site: heme iron (His, Met) (axial ligands) #status predicted

  Query Match 100.0%; Score 21; DB 2; Length 80;
  Best Local Similarity 100.0%; Pred. No. 2.4e+02;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDA 4
      ||||
Db 33 RGDA 36

RESULT 14
H82662
conserved hypothetical protein XF1562 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: H82662
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82662
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-88 <SIM>
A;Cross-references: GB:AE003986; GB:AE003849; NID:g9106606; PIDN:AAF84371.1; GSPDB:GN00128; XFSC:XF1562
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado,

J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1562

  Query Match   100.0%; Score 21; DB 2; Length 88;
  Best Local Similarity   100.0%; Pred. No. 2.7e+02;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   65 RGDA 68

RESULT 15
I68553
cell surface glycoprotein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 23-Jul-1999
C;Accession: I68553
R;Horn, G.T.; Bugawan, T.L.; Long, C.M.; Manos, M.M.; Erlich, H.A.
Hum. Immunol. 21, 249-263, 1988
A;Title: Sequence analysis of HLA class II genes from insulin-dependent diabetic individuals.
A;Reference number: I54290; MUID:88227495; PMID:3372263
A;Accession: I68553
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-89 <RES>
A;Cross-references: GB:M35000; NID:g291960; PIDN:AAA35774.1; PID:g553265
C;Superfamily: class II histocompatibility antigen; immunoglobulin homology
C;Keywords: glycoprotein

  Query Match   100.0%; Score 21; DB 2; Length 89;
  Best Local Similarity   100.0%; Pred. No. 2.7e+02;
  Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 RGDA 4
              ||||
Db   44 RGDA 47

Search completed: February 11, 2004, 14:56:56

Job time : 4.70968 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:36:52 ; Search time 1.67742 Seconds
(without alignments)
112.141 Million cell updates/sec

Title: US-10-050-611-1
Perfect score: 21
Sequence: 1 RGDA 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	39	1	DECO_MACDE	P17350 macrobdella
2	21	100.0	52	1	ORNC_PLAOR	P25512 placobdella
3	21	100.0	74	1	R38B_SCHPO	Q09900 schizosacch
4	21	100.0	76	1	MTRB_BACPU	P48064 bacillus pu
5	21	100.0	80	1	C551_CHRVI	P80549 chromatium
6	21	100.0	97	1	RL21_PYRAB	Q9uzp1 pyrococcus
7	21	100.0	97	1	RL21_PYRHO	O74001 pyrococcus
8	21	100.0	98	1	UL19_HCMVA	P16723 human cytom
9	21	100.0	113	1	APG1_HUMAN	Q15772 homo sapien
10	21	100.0	113	1	APG1_MOUSE	Q62407 mus musculu
11	21	100.0	113	1	APG1_RAT	Q63638 rattus norv
12	21	100.0	116	1	RL17_HELPJ	Q9zjt6 helicobacte
13	21	100.0	116	1	RL17_HELPY	P56042 helicobacte
14	21	100.0	124	1	RL17_MYCPN	Q59547 mycoplasma
15	21	100.0	124	1	RS8E_HALN1	Q9hpe9 halobacteri
16	21	100.0	131	1	RL17_THEMA	Q9x1i1 thermotoga
17	21	100.0	133	1	GEPE_BACSU	O06717 bacillus su
18	21	100.0	140	1	CO8B_RAT	P55314 rattus norv
19	21	100.0	141	1	NIKR_METJA	Q57969 methanococc
20	21	100.0	143	1	IR09_HCMVA	P16807 human cytom
21	21	100.0	149	1	DUT_CORGL	Q8npa9 corynebacte
22	21	100.0	150	1	FLAG_METVO	O06640 methanococc
23	21	100.0	150	1	MOAE_HAEIN	P45308 haemophilus
24	21	100.0	151	1	CP2B_DROME	Q9nip6 drosophila
25	21	100.0	155	1	RR7_CUSEU	P46292 cuscuta eur
26	21	100.0	157	1	Y510_VIBCH	Q9kuk8 vibrio chol
27	21	100.0	164	1	RL15_HALMA	P12737 haloarcula
28	21	100.0	168	1	TPX_CHLTE	Q8ked5 chlorobium
29	21	100.0	172	1	LBD4_ARATH	Q9she9 arabidopsis
30	21	100.0	177	1	RL6_HALMA	P14135 haloarcula
31	21	100.0	179	1	YF36_PSEAE	Q9i3h7 pseudomonas
32	21	100.0	181	1	YG86_STRCO	Q9s266 streptomyce
33	21	100.0	185	1	RRF_BUCAI	P57328 buchnera ap
34	21	100.0	186	1	YCE7_DROME	O97067 drosophila
35	21	100.0	190	1	Y2H5_STRCO	P35925 streptomyce
36	21	100.0	192	1	TERD_ALCSP	P18781 alcaligenes
37	21	100.0	197	1	HAM1_PSEAE	Q9i6a8 pseudomonas
38	21	100.0	201	1	EFA4_HUMAN	P52798 homo sapien
39	21	100.0	201	1	SODE_ONCVO	Q07449 onchocerca
40	21	100.0	202	1	B3G1_MOUSE	Q9cw73 m galactosy
41	21	100.0	203	1	IDI_MYCTU	P72002 mycobacteri
42	21	100.0	206	1	EFA4_MOUSE	O08542 mus musculu
43	21	100.0	206	1	YMAB_BACSU	P50619 bacillus su
44	21	100.0	212	1	RB17_HUMAN	Q9h0t7 homo sapien
45	21	100.0	214	1	RADC_VIBVU	Q8ddy0 vibrio vuln

ALIGNMENTS

RESULT 1
DECO_MACDE
ID DECO_MACDE STANDARD; PRT; 39 AA.
AC P17350;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Decorsin.
OS Macrobdella decora (North American leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC Arynchobdellida; Hirudiniformes; Hirudinidae; Macrobdella.
OX NCBI_TaxID=6405;
RN [1]
RP SEQUENCE.
RX MEDLINE=90277628; PubMed=2351655;
RA Seymour J.L., Henzel W.J., Nevins B., Stults J.T., Lazarus R.A.;
RT "Decorsin. A potent glycoprotein IIb-IIIa antagonist and platelet
RT aggregation inhibitor from the leech Macrobdella decora.";
RL J. Biol. Chem. 265:10143-10147(1990).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=94278502; PubMed=8009227;
RA Krezel A.M., Wagner G., Seymour-Ulmer J., Lazarus R.A.;
RT "Structure of the RGD protein decorsin: conserved motif and distinct

```
RT   function in leech proteins that affect blood clotting.";
RL   Science 264:1944-1947(1994).
CC   -!- FUNCTION: INHIBITS FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS
CC       EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT BLOOD FROM
CC       CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF INGESTED BLOOD.
CC   -!- SIMILARITY: HIGH, TO P.ORNATA ORNATINS.
CC   -!- SIMILARITY: SOME, TO THE DISINTEGRIN FAMILY.
DR   PIR; A36453; A36453.
DR   PDB; 1DEC; 31-AUG-94.
KW   Blood coagulation; Platelet; Cell adhesion; 3D-structure.
FT   DOMAIN       27     38       HIGH AFFINITY BINDING DOMAIN (POTENTIAL).
FT   SITE         31     33       CELL ATTACHMENT SITE.
FT   VARIANT       1      3       MISSING (IN N-3 ISOFORM).
FT   STRAND        6      6
FT   STRAND       15     16
FT   STRAND       21     22
FT   TURN         24     25
FT   STRAND       27     28
FT   STRAND       37     39
SQ   SEQUENCE   39 AA;  4384 MW;  3A3B35756FB70D36 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         31 RGDA 34


RESULT 2
ORNC_PLAOR
ID   ORNC_PLAOR     STANDARD;      PRT;    52 AA.
AC   P25512;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ornatin C.
OS   Placobdella ornata (Turtle leech).
OC   Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC   Rhynchobdellida; Glossiphoniidae; Placobdella.
OX   NCBI_TaxID=6415;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=92111479; PubMed=1765068;
RA   Mazur P., Henzel W.J., Seymour J.L., Lazarus R.A.;
RT   "Ornatins: potent glycoprotein IIb-IIIa antagonists and platelet
RT   aggregation inhibitors from the leech Placobdella ornata.";
RL   Eur. J. Biochem. 202:1073-1082(1991).
CC   -!- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET
CC       RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT
CC       BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF
CC       INGESTED BLOOD.
CC   -!- SIMILARITY: BELONGS TO THE ORNATIN FAMILY.
DR   PIR; S19623; S19623.
DR   InterPro; IPR002463; Ornatin.
DR   Pfam; PF02088; Ornatin; 1.
DR   ProDom; PD012062; Ornatin; 1.
KW   Blood coagulation; Platelet; Cell adhesion.
FT   SITE         42     44       CELL ATTACHMENT SITE.
SQ   SEQUENCE   52 AA;  5845 MW;  BA55CA7408EF4F09 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         42 RGDA 45


RESULT 3
R38B_SCHPO
ID   R38B_SCHPO     STANDARD;      PRT;    74 AA.
AC   Q09900;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60S ribosomal protein L38-2.
GN   RPL38B OR RPL38 OR SPAC30D11.12.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
```

```
RL   Nature 415:871-880(2002).
CC   -!- MISCELLANEOUS: There are two genes for L38 in S.pombe.
CC   -!- SIMILARITY: BELONGS TO THE L38E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z67961; CAA91898.1; -.
DR   PIR; S62570; S62570.
DR   GeneDB_SPombe; SPAC30D11.12; -.
DR   InterPro; IPR002675; Ribosomal_L38e.
DR   Pfam; PF01781; Ribosomal_L38e; 1.
DR   ProDom; PD010361; Ribosomal_L38e; 1.
KW   Ribosomal protein; Multigene family.
SQ   SEQUENCE   74 AA;  8339 MW;  C90D6594DFCB11D3 CRC64;

  Query Match           100.0%;  Score 21;  DB 1;  Length 74;
  Best Local Similarity  100.0%;  Pred. No. 94;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         17 RGDA 20


RESULT 4
MTRB_BACPU
ID   MTRB_BACPU      STANDARD;      PRT;    76 AA.
AC   P48064;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription attenuation protein mtrB (Tryptophan RNA-binding
DE   attenuator protein) (Trp RNA-binding attenuation protein) (TRAP).
GN   MTRB.
OS   Bacillus pumilus (Bacillus mesentericus).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1408;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95138053; PubMed=7836324;
RA   Hoffman R.J., Gollnick P.;
RT   "The mtrB gene of Bacillus pumilus encodes a protein with sequence
RT   and functional homology to the trp RNA-binding attenuation protein
RT   (TRAP) of Bacillus subtilis.";
RL   J. Bacteriol. 177:839-842(1995).
CC   -!- FUNCTION: REQUIRED FOR TRANSCRIPTION ATTENUATION CONTROL IN THE
CC       TRP OPERON. THIS TRANS-ACTING FACTOR SEEMS TO RECOGNIZE A 10 BASES
CC       NUCLEOTIDE SEQUENCE IN THE TRP LEADER TRANSCRIPT CAUSING
CC       TRANSCRIPTION TERMINATION. BINDS THE LEADER RNA ONLY IN PRESENCE
CC       OF L-TRYPTOPHAN.
CC   -!- SUBUNIT: OLIGOMER OF 11 IDENTICAL SUBUNITS ARRANGED IN DOUGHNUT-
CC       LIKE STRUCTURE (BY SIMILARITY).
CC   -!- SIMILARITY: WITH REGA FROM PHAGE T4.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L37879; AAA67544.1; -.
DR   PIR; I39905; I39905.
DR   HSSP; Q9X6J6; 1QAW.
DR   InterPro; IPR000824; TrpBP.
DR   Pfam; PF02081; TrpBP; 1.
DR   PRINTS; PR00687; TRPRNAAP.
DR   ProDom; PD027918; TrpBP; 1.
KW   Transcription regulation; RNA-binding.
SQ   SEQUENCE   76 AA;  8301 MW;  22184B2351DA151D CRC64;

  Query Match           100.0%;  Score 21;  DB 1;  Length 76;
  Best Local Similarity  100.0%;  Pred. No. 97;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         58 RGDA 61


RESULT 5
C551_CHRVI
ID   C551_CHRVI      STANDARD;      PRT;    80 AA.
AC   P80549;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Cytochrome c-551 (C551).
OS   Chromatium vinosum.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;
OC   Chromatiaceae; Allochromatium.
OX   NCBI_TaxID=1049;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=D / ATCC 17899 / DSM 180;
RX   MEDLINE=96195682; PubMed=8612646;
RA   Samyn B., de Smet L., van Driessche G., Meyer T.E., Bartsch R.G.,
RA   Cusanovich M.A., van Beeumen J.J.;
RT   "A high-potential soluble cytochrome c-551 from the purple
RT   phototrophic bacterium Chromatium vinosum is homologous to cytochrome
RT   c8 from denitrifying pseudomonads.";
RL   Eur. J. Biochem. 236:689-696(1996).
CC   -!- FUNCTION: MONOHEME CYTOCHROME.
DR   PIR; S68677; S68677.
DR   HSSP; P95339; 1A56.
DR   InterPro; IPR003088; Cyt_CI.
DR   InterPro; IPR002324; Cyt_CID.
```

```
DR   InterPro; IPR000345; CytC_heme_bind.
DR   Pfam; PF00034; cytochrome_c; 1.
DR   PRINTS; PR00606; CYTCHROMECID.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Electron transport; Heme.
FT   BINDING      10     10       HEME (COVALENT).
FT   BINDING      13     13       HEME (COVALENT).
FT   METAL        14     14       IRON (HEME AXIAL LIGAND).
FT   METAL        59     59       IRON (HEME AXIAL LIGAND).
SQ   SEQUENCE   80 AA;  8224 MW;  EBD30A2B15D07F93 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         33 RGDA 36


RESULT 6
RL21_PYRAB
ID   RL21_PYRAB     STANDARD;      PRT;    97 AA.
AC   Q9UZP1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   50S ribosomal protein L21e.
GN   RPL21E OR PYRAB11050 OR PAB0731.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GE5 / Orsay;
RX   PubMed=12622808;
RA   Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA   Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA   Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT   "An integrated analysis of the genome of the hyperthermophilic
RT   archaeon Pyrococcus abyssi.";
RL   Mol. Microbiol. 47:1495-1512(2003).
CC   -!- SIMILARITY: BELONGS TO THE L21E FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ248286; CAB50016.1; -.
DR   PIR; C75089; C75089.
DR   HAMAP; MF_00369; -; 1.
DR   InterPro; IPR001147; Ribosomal_L21e.
DR   Pfam; PF01157; Ribosomal_L21e; 1.
DR   PROSITE; PS01171; RIBOSOMAL_L21E; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   97 AA;  11378 MW;  6CEF3A2DB6A61E40 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 97;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         69 RGDA 72


RESULT 7
RL21_PYRHO
ID   RL21_PYRHO     STANDARD;      PRT;    97 AA.
AC   O74001;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   50S ribosomal protein L21e.
GN   RPL21E OR PH1127.1 OR PHS032.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- SIMILARITY: BELONGS TO THE L21E FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP000005; BAA30227.1; -.
DR   PIR; A71054; A71054.
DR   HAMAP; MF_00369; -; 1.
DR   InterPro; IPR001147; Ribosomal_L21e.
DR   Pfam; PF01157; Ribosomal_L21e; 1.
DR   PROSITE; PS01171; RIBOSOMAL_L21E; 1.
KW   Ribosomal protein; Complete proteome.
```

```
SQ   SEQUENCE   97 AA;  11376 MW;  6D5D229DBFBE0E51 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 97;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         69 RGDA 72


RESULT 8
UL19_HCMVA
ID   UL19_HCMVA     STANDARD;      PRT;    98 AA.
AC   P16723;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   Hypothetical protein UL19.
GN   UL19.
OS   Human cytomegalovirus (strain AD169).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=10360;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88094735; PubMed=2827039;
RA   Beck S., Barrell B.G.;
RT   "Human cytomegalovirus encodes a glycoprotein homologous to MHC
RT   class-I antigens.";
RL   Nature 331:269-272(1988).
RN   [2]
RP   COMPLETE GENOME.
RX   MEDLINE=90269039; PubMed=2161319;
RA   Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA   Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA   Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT   "Analysis of the protein-coding content of the sequence of human
RT   cytomegalovirus strain AD169.";
RL   Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Y00293; -; NOT_ANNOTATED_CDS.
DR   EMBL; X17403; CAA35418.1; -.
DR   PIR; S01566; S01566.
KW   Hypothetical protein.
SQ   SEQUENCE   98 AA;  11280 MW;  7E8A7405611E3F2B CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 98;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         95 RGDA 98


RESULT 9
APG1_HUMAN
ID   APG1_HUMAN     STANDARD;      PRT;   113 AA.
AC   Q15772;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Aortic preferentially expressed protein 1 (APEG-1).
GN   APEG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96291890; PubMed=8663449;
RA   Hsieh C.-M., Yoshizumi M., Endege W.O., Kho C.-J., Jain M.K.,
RA   Kashiki S., de Los Santos R., Lee W.-S., Perrella M.A., Lee M.-E.;
RT   "APEG-1, a novel gene preferentially expressed in aortic smooth muscle
RT   cells, is down-regulated by vascular injury.";
RL   J. Biol. Chem. 271:17354-17359(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MAY HAVE A ROLE IN REGULATING THE GROWTH AND
CC       DIFFERENTIATION OF ARTERIAL SMOOTH MUSCLE CELLS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN DIFFERENTIATED
CC       ARTERIAL SMOOTH MUSCLE CELLS (ASMC).
```

```
CC   -!- DEVELOPMENTAL STAGE: APPEARS TO BE EXPRESSED ONLY IN HIGHLY
CC       DIFFERENTIATED ASMC IN NORMAL VESSEL WALLS AND DOWN-REGULATED IN
CC       DEDIFFERENTIATED ASMC IN VIVO. IN RESPONSE TO VASCULAR INJURIES
CC       ASMC DEDIFFERENTIATE AND CHANGE FROM A QUIESCENT AND CONTRACTILE
CC       PHENOTYPE TO A PROLIFERATIVE AND SYNTHETIC PHENOTYPE. THIS
CC       PROLIFERATION OF VASCULAR SMOTH MUSCLE CELLS IS ONE OF THE MOST
CC       PROMINENT FEATURES OF ARTIOSCLEROSIS.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U57099; AAC50599.1; -.
DR   EMBL; BC006346; AAH06346.1; -.
DR   HSSP; P56276; 1TLK.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0007517; P:muscle development; TAS.
DR   GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin domain; Nuclear protein.
FT   DOMAIN       20    109       IG-LIKE.
SQ   SEQUENCE   113 AA;  12692 MW;  04F367263A1397C5 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         85 RGDA 88


RESULT 10
APG1_MOUSE
ID   APG1_MOUSE     STANDARD;      PRT;   113 AA.
AC   Q62407;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Aortic preferentially expressed protein 1 (APEG-1).
GN   APEG1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=96291890; PubMed=8663449;
RA   Hsieh C.-M., Yoshizumi M., Endege W.O., Kho C.-J., Jain M.K.,
RA   Kashiki S., de Los Santos R., Lee W.-S., Perrella M.A., Lee M.-E.;
RT   "APEG-1, a novel gene preferentially expressed in aortic smooth muscle
RT   cells, is down-regulated by vascular injury.";
RL   J. Biol. Chem. 271:17354-17359(1996).
CC   -!- FUNCTION: MAY HAVE A ROLE IN REGULATING THE GROWTH AND
CC       DIFFERENTIATION OF ARTERIAL SMOOTH MUSCLE CELLS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN DIFFERENTIATED
CC       ARTERIAL SMOOTH MUSCLE CELLS (ASMC).
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U57098; AAC52666.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:109282; Apeg1.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin domain; Nuclear protein.
FT   DOMAIN       20    109       IG-LIKE.
SQ   SEQUENCE   113 AA;  12665 MW;  5F320C5A41C3D870 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         85 RGDA 88


RESULT 11
APG1_RAT
ID   APG1_RAT       STANDARD;      PRT;   113 AA.
AC   Q63638;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Aortic preferentially expressed protein 1 (APEG-1).
GN   APEG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=96291890; PubMed=8663449;
RA   Hsieh C.-M., Yoshizumi M., Endege W.O., Kho C.-J., Jain M.K.,
RA   Kashiki S., de Los Santos R., Lee W.-S., Perrella M.A., Lee M.-E.;
RT   "APEG-1, a novel gene preferentially expressed in aortic smooth muscle
RT   cells, is down-regulated by vascular injury.";
RL   J. Biol. Chem. 271:17354-17359(1996).
CC   -!- FUNCTION: MAY HAVE A ROLE IN REGULATING THE GROWTH AND
CC       DIFFERENTIATION OF ARTERIAL SMOOTH MUSCLE CELLS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN DIFFERENTIATED ARTERIAL
CC       SMOOTH MUSCLE CELLS (ASMC) IN THE MEDIAL LAYER OF THE AORTA.
CC       WEAKLY DETECTED IN BRAIN AND TESTIS AND TO A LESSER EXTENT IN
CC       ORGANS RICH IN STRIATED MUSCLE OR VISCERAL SMOOTH MUSCLE.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U57097; AAC52667.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin domain; Nuclear protein.
FT   DOMAIN       20    109       IG-LIKE.
SQ   SEQUENCE   113 AA;  12668 MW;  B213C366A759A363 CRC64;

  Query Match            100.0%;  Score 21;  DB 1;  Length 113;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         85 RGDA 88


RESULT 12
RL17_HELPJ
ID   RL17_HELPJ     STANDARD;      PRT;   116 AA.
AC   Q9ZJT6;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   50S ribosomal protein L17.
GN   RPLQ OR JHP1212.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001547; AAD06814.1; -.
DR   PIR; D71832; D71832.
DR   InterPro; IPR000456; Ribosomal_L17.
DR   Pfam; PF01196; Ribosomal_L17; 1.
DR   ProDom; PD004277; Ribosomal_L17; 1.
DR   TIGRFAMs; TIGR00059; L17; 1.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   116 AA;  13392 MW;  EBC77780E2F2F3A1 CRC64;

  Query Match            100.0%;  Score 21;  DB 1;  Length 116;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db        104 RGDA 107


RESULT 13
RL17_HELPY
ID   RL17_HELPY     STANDARD;      PRT;   116 AA.
AC   P56042;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   50S ribosomal protein L17.
GN   RPLQ OR HP1292.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000633; AAD08335.1; -.
DR   PIR; D64681; D64681.
DR   TIGR; HP1292; -.
DR   InterPro; IPR000456; Ribosomal_L17.
DR   Pfam; PF01196; Ribosomal_L17; 1.
DR   ProDom; PD004277; Ribosomal_L17; 1.
DR   TIGRFAMs; TIGR00059; L17; 1.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   116 AA;  13364 MW;  EBD87890E2F2E4B6 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db        104 RGDA 107


RESULT 14
RL17_MYCPN
ID   RL17_MYCPN     STANDARD;      PRT;   124 AA.
AC   Q59547;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   50S ribosomal protein L17.
GN   RPLQ OR MPN192 OR MP639.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=96177562; PubMed=8604303;
RA   Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
RT   "Sequence analysis of 56 kb from the genome of the bacterium
RT   Mycoplasma pneumoniae comprising the dnaA region, the atp operon and
RT   a cluster of ribosomal protein genes.";
RL   Nucleic Acids Res. 24:628-639(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34795; AAC43689.1; -.
DR   EMBL; AE000061; AAB96287.1; -.
DR   PIR; S62816; S62816.
DR   InterPro; IPR000456; Ribosomal_L17.
DR   Pfam; PF01196; Ribosomal_L17; 1.
DR   TIGRFAMs; TIGR00059; L17; 1.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   124 AA;  14245 MW;  3A627DB7EBF8C62E CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db        107 RGDA 110


RESULT 15
RS8E_HALN1
ID   RS8E_HALN1     STANDARD;      PRT;   124 AA.
AC   Q9HPE9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S8e.
GN   RPS8E OR VNG1668G.
OS   Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
```

```
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- SIMILARITY: BELONGS TO THE S8E FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE005076; AAG19920.1; -.
DR   PIR; D84319; D84319.
DR   HAMAP; MF_00029; -; 1.
DR   InterPro; IPR001047; Ribosomal_S8E.
DR   Pfam; PF01201; Ribosomal_S8e; 1.
DR   ProDom; PD005658; Ribosomal_S8E; 1.
DR   TIGRFAMs; TIGR00307; S8e; 1.
DR   PROSITE; PS01193; RIBOSOMAL_S8E; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   124 AA;  13515 MW;  B7038CF79A83742B CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         47 RGDA 50


Search completed: February 11, 2004, 14:54:03
Job time : 4.67742 secs
```

```
                                GenCore version 5.1.6
                          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          February 11, 2004, 14:47:57 ; Search time 6.83871 Seconds
                                             (without alignments)
                                             150.936 Million cell updates/sec

Title:           US-10-050-611-1
Perfect score:   21
Sequence:        1 RGDA 4

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        SPTREMBL_23:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                    SUMMARIES

                          %
Result                  Query
  No.    Score          Match  Length DB  ID                      Description
-------------------------------------------------------------------------------
```

1	21	100.0	31	5	Q8MXE8	Q8mxe8 caenorhabdi
2	21	100.0	45	16	Q9PGB6	Q9pgb6 xylella fas
3	21	100.0	48	2	Q9XDV3	Q9xdv3 erythrobact
4	21	100.0	54	16	Q8R7H3	Q8r7h3 thermoanaer
5	21	100.0	55	10	Q8RUZ1	Q8ruz1 zea mays (m
6	21	100.0	57	6	Q9N041	Q9n041 macaca fasc
7	21	100.0	57	10	Q8RUD5	Q8rud5 zea mays (m
8	21	100.0	57	10	Q8RUD4	Q8rud4 zea mays (m
9	21	100.0	57	16	O06773	O06773 mycobacteri
10	21	100.0	58	12	Q8QS83	Q8qs83 chimpanzee
11	21	100.0	59	16	Q98LS7	Q98ls7 rhizobium l
12	21	100.0	64	16	Q8XYQ0	Q8xyq0 ralstonia s
13	21	100.0	66	12	Q8JKZ2	Q8jkz2 virus phich
14	21	100.0	68	5	Q8MNA5	Q8mna5 dictyosteli
15	21	100.0	68	16	Q8UJK6	Q8ujk6 agrobacteri
16	21	100.0	69	16	Q8DDL7	Q8ddl7 vibrio vuln
17	21	100.0	70	12	Q8VAV0	Q8vav0 white spot
18	21	100.0	70	16	Q8XTW3	Q8xtw3 ralstonia s
19	21	100.0	73	16	Q8Y128	Q8y128 ralstonia s
20	21	100.0	75	16	Q8VJ45	Q8vj45 mycobacteri
21	21	100.0	76	10	Q8GVK2	Q8gvk2 oryza sativ
22	21	100.0	77	6	Q29171	Q29171 sus scrofa
23	21	100.0	77	16	Q92KI0	Q92ki0 rhizobium m
24	21	100.0	79	16	Q8X8Q7	Q8x8q7 escherichia
25	21	100.0	83	17	Q8TK40	Q8tk40 methanosarc
26	21	100.0	85	10	Q8W3B8	Q8w3b8 oryza sativ
27	21	100.0	88	16	Q9PD18	Q9pd18 xylella fas
28	21	100.0	88	17	Q8ZV78	Q8zv78 pyrobaculum
29	21	100.0	89	5	Q95Y01	Q95y01 caenorhabdi
30	21	100.0	89	7	Q29783	Q29783 homo sapien
31	21	100.0	89	16	Q8G801	Q8g801 bifidobacte
32	21	100.0	90	16	Q9PAU0	Q9pau0 xylella fas
33	21	100.0	91	15	Q9DK41	Q9dk41 human immun
34	21	100.0	91	16	Q8PJH2	Q8pjh2 xanthomonas
35	21	100.0	92	9	Q9FZT5	Q9fzt5 pseudomonas
36	21	100.0	93	10	Q8S2D8	Q8s2d8 oryza sativ
37	21	100.0	93	16	Q8Z7W3	Q8z7w3 salmonella
38	21	100.0	95	16	Q9PDS1	Q9pds1 xylella fas
39	21	100.0	95	16	Q98FE9	Q98fe9 rhizobium l
40	21	100.0	96	16	Q9KE84	Q9ke84 bacillus ha
41	21	100.0	96	17	Q9HR67	Q9hr67 halobacteri
42	21	100.0	97	16	Q9HTA8	Q9hta8 pseudomonas
43	21	100.0	99	2	Q8RM68	Q8rm68 bacteroides
44	21	100.0	100	12	Q98239	Q98239 molluscum c
45	21	100.0	100	12	Q8B9W5	Q8b9w5 influenza b

ALIGNMENTS

```
RESULT 1
Q8MXE8
ID   Q8MXE8      PRELIMINARY;      PRT;    31 AA.
AC   Q8MXE8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein K07A9.4.
GN   K07A9.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Davidson S., O'Neal D.;
RT   "The sequence of C. elegans cosmid K07A9.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF099924; AAM98005.1; -.
DR   WormPep; K07A9.4; CE31709.
KW   Hypothetical protein.
SQ   SEQUENCE   31 AA;  3720 MW;  147938913DC940ED CRC64;

  Query Match           100.0%;  Score 21;  DB 5;  Length 31;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          2 RGDA 5


RESULT 2
Q9PGB6
ID   Q9PGB6      PRELIMINARY;      PRT;    45 AA.
AC   Q9PGB6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Xf0386.
GN   XF0386.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
```

```
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003890; AAF83196.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   45 AA;  5163 MW;  B58C9AECC9809C8A CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         19 RGDA 22


RESULT 3
Q9XDV3
ID   Q9XDV3      PRELIMINARY;      PRT;    48 AA.
AC   Q9XDV3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ORF Q.
OS   Erythrobacter sp. MBIC3960.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC   Sphingomonadaceae; Erythrobacter.
OX   NCBI_TaxID=94771;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MBIC3960;
RA   Hamada T.;
RT   "Nucleotide sequences of genes coding for photosynthetic reaction
RT   centers and light-harvesting proteins of Erythrobacter litoralis and
RT   related aerobic photosynthetic bacteria.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB027515; BAA78669.1; -.
DR   InterPro; IPR006089; Acyl-CoA_dh.
DR   PROSITE; PS00073; ACYL_COA_DH_2; 1.
SQ   SEQUENCE   48 AA;  4980 MW;  D663EAD05EA8079B CRC64;

  Query Match             100.0%;  Score 21;  DB 2;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         27 RGDA 30


RESULT 4
Q8R7H3
ID   Q8R7H3      PRELIMINARY;      PRT;    54 AA.
AC   Q8R7H3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein TTE2436.
GN   TTE2436.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC   Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4 / JCM 11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013185; AAM25571.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   54 AA;  6252 MW;  0A9C818C07DD905B CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         32 RGDA 35


RESULT 5
Q8RUZ1
ID   Q8RUZ1      PRELIMINARY;      PRT;    55 AA.
AC   Q8RUZ1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Acetyl-CoA C-acyltransferase-like protein (Fragment).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Various strains;
RA   Ching A.S., Caldwell K.S., Jung M., Dolan M., Smith O.S., Tingey S.,
RA   Morgante M., Rafalski J.A.;
RT   "SNP frequency, haplotype structure and linkage disequilibrium in
RT   elite maize inbred lines.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF498463; AAM14479.1; -.
DR   EMBL; AF498469; AAM14485.1; -.
DR   EMBL; AF498472; AAM14488.1; -.
DR   EMBL; AF498477; AAM14493.1; -.
DR   EMBL; AF498482; AAM14498.1; -.
DR   EMBL; AF498485; AAM14501.1; -.
DR   EMBL; AF498486; AAM14502.1; -.
DR   InterPro; IPR002155; Thiolase.
DR   Pfam; PF02803; thiolase_C; 1.
DR   PROSITE; PS00099; THIOLASE_3; 1.
KW   Acyltransferase; Transferase.
FT   NON_TER       1      1
SQ   SEQUENCE   55 AA;  5959 MW;  5C09DAC7224451D0 CRC64;

  Query Match             100.0%;  Score 21;  DB 10;  Length 55;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         31 RGDA 34


RESULT 6
Q9N041
ID   Q9N041      PRELIMINARY;      PRT;    57 AA.
AC   Q9N041;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   Unnamed protein product.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB046091; BAB01673.1; -.
SQ   SEQUENCE   57 AA;  6250 MW;  300DE0464A4897A9 CRC64;

  Query Match             100.0%;  Score 21;  DB 6;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 5.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         10 RGDA 13


RESULT 7
Q8RUD5
ID   Q8RUD5      PRELIMINARY;      PRT;    57 AA.
AC   Q8RUD5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Acetyl-CoA C-acyltransferase-like protein (Fragment).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Various strains;
RA   Ching A.S., Caldwell K.S., Jung M., Dolan M., Smith O.S., Tingey S.,
RA   Morgante M., Rafalski J.A.;
RT   "SNP frequency, haplotype structure and linkage disequilibrium in
RT   elite maize inbred lines.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF498457; AAM14473.1; -.
DR   EMBL; AF498458; AAM14474.1; -.
DR   EMBL; AF498459; AAM14475.1; -.
DR   EMBL; AF498460; AAM14476.1; -.
DR   EMBL; AF498461; AAM14477.1; -.
DR   EMBL; AF498462; AAM14478.1; -.
DR   EMBL; AF498464; AAM14480.1; -.
DR   EMBL; AF498465; AAM14481.1; -.
DR   EMBL; AF498466; AAM14482.1; -.
DR   EMBL; AF498467; AAM14483.1; -.
DR   EMBL; AF498468; AAM14484.1; -.
DR   EMBL; AF498470; AAM14486.1; -.
DR   EMBL; AF498471; AAM14487.1; -.
DR   EMBL; AF498473; AAM14489.1; -.
DR   EMBL; AF498475; AAM14491.1; -.
DR   EMBL; AF498478; AAM14494.1; -.
DR   EMBL; AF498480; AAM14496.1; -.
DR   EMBL; AF498481; AAM14497.1; -.
DR   EMBL; AF498483; AAM14499.1; -.
DR   EMBL; AF498484; AAM14500.1; -.
DR   EMBL; AF498487; AAM14503.1; -.
DR   InterPro; IPR002155; Thiolase.
```

```
DR   Pfam; PF02803; thiolase_C; 1.
DR   PROSITE; PS00099; THIOLASE_3; 1.
KW   Acyltransferase; Transferase.
FT   NON_TER       1      1
SQ   SEQUENCE   57 AA;  6203 MW;  DC4596C27A4451A8 CRC64;

  Query Match             100.0%;  Score 21;  DB 10;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 5.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         33 RGDA 36


RESULT 8
Q8RUD4
ID   Q8RUD4      PRELIMINARY;      PRT;    57 AA.
AC   Q8RUD4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Acetyl-CoA C-acyltransferase-like protein (Fragment).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. IVANA, cv. D71-4HT, and cv. H60;
RA   Ching A.S., Caldwell K.S., Jung M., Dolan M., Smith O.S., Tingey S.,
RA   Morgante M., Rafalski J.A.;
RT   "SNP frequency, haplotype structure and linkage disequilibrium in
RT   elite maize inbred lines.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF498474; AAM14490.1; -.
DR   EMBL; AF498476; AAM14492.1; -.
DR   EMBL; AF498479; AAM14495.1; -.
DR   InterPro; IPR002155; Thiolase.
DR   Pfam; PF02803; thiolase_C; 1.
DR   PROSITE; PS00099; THIOLASE_3; 1.
KW   Acyltransferase; Transferase.
FT   NON_TER       1      1
SQ   SEQUENCE   57 AA;  6185 MW;  DC4596C76E4451A8 CRC64;

  Query Match             100.0%;  Score 21;  DB 10;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 5.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         33 RGDA 36


RESULT 9
O06773
ID   O06773      PRELIMINARY;      PRT;    57 AA.
AC   O06773;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Rv0666.
GN   RV0666 OR MTCI376.10C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; Z95972; CAB09391.1; -.
DR   TubercuList; Rv0666; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   57 AA;  5849 MW;  62858455BD7D0F2E CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 5.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         24 RGDA 27


RESULT 10
Q8QS83
ID   Q8QS83      PRELIMINARY;      PRT;    58 AA.
AC   Q8QS83;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   UL2.
OS   Chimpanzee cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=188763;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Davison A.J., Akter P., Dolan A., Wright K.M., Addison C.,
RA   Alcendor D.J., Hayward G.S., McGeoch D.J.;
RT   "The human cytomegalovirus genome revisited.";
```

```
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF480884; AAM00654.1; -.
SQ   SEQUENCE   58 AA;  6789 MW;  27400659BBD2BAD7 CRC64;

  Query Match             100.0%;  Score 21;  DB 12;  Length 58;
  Best Local Similarity   100.0%;  Pred. No. 5.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db          2 RGDA 5


RESULT 11
Q98LS7
ID   Q98LS7      PRELIMINARY;      PRT;    59 AA.
AC   Q98LS7;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein msl0897.
GN   MSL0897.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP002996; BAB48386.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   59 AA;  6059 MW;  4EE77EF3940E6633 CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 5.6e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         36 RGDA 39


RESULT 12
Q8XYQ0
ID   Q8XYQ0      PRELIMINARY;      PRT;    64 AA.
AC   Q8XYQ0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein RSc1708.
GN   RSC1708 OR RS02894.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Ralstoniaceae; Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646066; CAD15410.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   64 AA;  7210 MW;  F35F8AEF5E609609 CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 6.1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         60 RGDA 63


RESULT 13
Q8JKZ2
ID   Q8JKZ2      PRELIMINARY;      PRT;    66 AA.
AC   Q8JKZ2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Virus PhiCh1.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX   NCBI_TaxID=114777;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20177831; PubMed=10712697;
RA   Baranyi U., Klein R., Lubitz W., Kruger D.H., Witte A.;
RT   "The archaeal halophilic virus-encoded Dam-like methyltransferase M.
RT   phiCh1-I methylates adenine residues and complements dam mutants in
RT   the low salt environment of Escherichia coli.";
RL   Mol. Microbiol. 35:1168-1179(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20497008; PubMed=11040128;
RA   Klein R., Greineder B., Baranyi U., Witte A.;
RT   "The structural protein E of the archaeal virus phiCh1: evidence for
RT   processing in Natrialba magadii during virus maturation.";
```

```
RL   Virology 276:376-387(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22136043; PubMed=12139629;
RA   Klein R., Baranyi U., Rossler N., Greineder B., Scholz H., Witte A.;
RT   "Natrialba magadii virus phiCh1: first complete nucleotide sequence
RT   and functional organization of a virus infecting a haloalkaliphilic
RT   archaeon.";
RL   Mol. Microbiol. 45:851-863(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Klein R., Baranyi U., Roessler N., Greineder B., Scholz H.;
RT   "Sequence analysis of the temperate virus PhiCh1 infecting the
RT   haloalkalophilic archaeon Natrialba magadii.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF440695; AAM88738.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   66 AA;  6695 MW;  38EA1246C5F281A6 CRC64;

  Query Match             100.0%;  Score 21;  DB 12;  Length 66;
  Best Local Similarity   100.0%;  Pred. No. 6.3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGDA 4
              ||||
Db         20 RGDA 23


RESULT 14
Q8MNA5
ID   Q8MNA5      PRELIMINARY;      PRT;    68 AA.
AC   Q8MNA5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117076; AAM33713.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   68 AA;  7790 MW;  C2E2D3DA9412A754 CRC64;

  Query Match             100.0%;  Score 21;  DB 5;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
Qy          1 RGDA 4
              ||||
Db         41 RGDA 44


RESULT 15
Q8UJK6
ID   Q8UJK6      PRELIMINARY;      PRT;    68 AA.
AC   Q8UJK6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Atu5470.
GN   ATU5470 OR AGR_PAT_693.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OG   Plasmid AT.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE008968; AAL46157.1; -.
DR   EMBL; AE007916; AAK90845.1; -.
KW   Hypothetical protein; Plasmid; Complete proteome.
SQ   SEQUENCE   68 AA;  8005 MW;  5CABE406D75F93A8 CRC64;
  Query Match             100.0%;  Score 21;  DB 16;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
Qy          1 RGDA 4
              ||||
Db         36 RGDA 39

Search completed: February 11, 2004, 14:56:02
Job time : 9.83871 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:35:52 ; Search time 25.9355 Seconds
(without alignments)
73.441 Million cell updates/sec

Title: US-10-050-611-2
Perfect score: 69
Sequence: 1 DACEGDSGGPFV 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	100.0	12	23	AAM50857	Serine esterase co
2	69	100.0	23	20	AAW83414	Cell growth/adhesi
3	69	100.0	23	21	AAB12893	Nerve tissue regen
4	69	100.0	23	22	AAB70363	Human thrombin rec
5	69	100.0	23	23	AAE22563	Human thrombin hig
6	69	100.0	23	23	AAE20159	Human thrombin pep
7	69	100.0	23	23	AAU78376	Thrombin peptide d
8	69	100.0	23	23	AAM50858	Thrombin-derived p
9	69	100.0	23	24	ABP72755	Antiulcer peptide
10	69	100.0	23	24	ABP72757	Antiulcer peptide
11	69	100.0	23	24	ABP72760	Human thrombin pep
12	69	100.0	33	24	ABP72758	Antiulcer peptide
13	69	100.0	111	20	AAW99113	Bovine zeta 2 pret
14	69	100.0	116	20	AAW99115	Human zeta 2 preth
15	69	100.0	259	18	AAW11545	Human thrombin Asn
16	69	100.0	259	24	ABP60563	Human thrombin var
17	69	100.0	259	24	ABP60565	Human thrombin var
18	69	100.0	295	16	AAR74775	Wild-type thrombin
19	69	100.0	295	16	AAR74776	Mutant thrombin K5
20	69	100.0	295	16	AAR74777	Mutant thrombin E2
21	69	100.0	295	16	AAR74778	Mutant thrombin E2
22	69	100.0	295	16	AAR74779	Mutant thrombin E2
23	69	100.0	295	16	AAR74780	Mutant thrombin E2
24	69	100.0	295	16	AAR76033	Mutant thrombin E2
25	69	100.0	295	16	AAR76034	Mutant thrombin R2
26	69	100.0	295	16	AAR76035	Mutant thrombin R2
27	69	100.0	295	16	AAR76036	Mutant thrombin R2
28	69	100.0	295	16	AAR76037	Mutant thrombin W5
29	69	100.0	295	16	AAR76038	Mutant thrombin K5
30	69	100.0	295	16	AAR76039	Mutant thrombin W5
31	69	100.0	295	16	AAR76040	Mutant thrombin W5
32	69	100.0	295	18	AAW22892	Human mature throm
33	69	100.0	295	21	AAB08633	Amino acid sequenc
34	69	100.0	295	24	ABP60562	Human thrombin var
35	69	100.0	295	24	ABP60564	Human thrombin var
36	69	100.0	308	20	AAW99107	Bovine prethrombin
37	69	100.0	308	20	AAW99109	Human prethrombin
38	69	100.0	376	14	AAR41797	CD4/Thrombin fusio
39	69	100.0	376	20	AAY42789	Human CD4/thrombin
40	69	100.0	376	23	AAU10703	Human CD4-thrombin
41	69	100.0	579	14	AAR35763	Prothrombin (PT).
42	69	100.0	579	18	AAW11546	Human prothrombin
43	69	100.0	579	18	AAW11544	Human prothrombin
44	69	100.0	579	20	AAW99108	Human prothrombin.
45	69	100.0	582	20	AAW99106	Bovine prothrombin

ALIGNMENTS

RESULT 1
AAM50857
ID AAM50857 standard; Peptide; 12 AA.
XX
AC AAM50857;
XX
DT 01-MAY-2002 (first entry)
XX
DE Serine esterase conserved sequence used in cardiac tissue repair.
XX
KW Serine esterase; thrombin; revascularisation; vascular occlusion;
KW tissue repair; vulnerary; vasotropic; cardiant; angiogenesis;
KW restenosis; therapy; enzyme; human.
XX
OS Homo sapiens.
XX
PN WO200204008-A2.
XX
PD 17-JAN-2002.
XX
PF 12-JUL-2001; 2001WO-US21944.
XX
PR 12-JUL-2000; 2000US-217583P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH;
XX
DR WPI; 2002-179665/23.
XX
PT Promoting cardiac tissue repair, stimulating revascularisation,
PT stimulating vascular endothelial cell proliferation, and inhibiting
PT vascular occlusion by using angiogenic thrombin derivative peptide -
XX
PS Claim 3; Page 19; 24pp; English.
XX
CC The present peptide comprises a thrombin-derived serine esterase
CC conserved sequence that is used in a claimed method for promoting
CC cardiac tissue repair. The method involves administering an
CC angiogenic thrombin-derived peptide, especially a thrombin receptor
CC binding domain comprising the 4-amino acid peptide given in
CC AAM50856 together with the serine esterase conserved sequence,
CC or preferably the peptide given in AAM50858, which includes both
CC these peptide sequences. The thrombin-derived peptide is
CC administered during or following cardiac surgery by injection
CC into cardiac tissue, and may be formulated as a sustained release
CC formulation. It is used in claimed methods of stimulating
CC revascularisation, stimulating vascular endothelial cell
CC proliferation, inhibiting vascular occlusion, and inhibiting
CC restenosis following balloon angioplasty, in which case the
CC peptide may be coated onto the catheter.
XX
SQ Sequence 12 AA;

Query Match 100.0%; Score 69; DB 23; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
    ||||||||||||
Db 1 DACEGDSGGPFV 12

RESULT 2
AAW83414
ID AAW83414 standard; peptide; 23 AA.
XX
AC AAW83414;
XX
DT 26-FEB-1999 (first entry)
XX
DE Cell growth/adhesion promoting peptide #1.
XX
KW Cell growth; adhesion; promotion; medical treatment; injury;
KW biotissue; bone reinforcement; nerve regeneration; HMP resin.
XX
OS Synthetic.
XX
PN JP10316581-A.
XX
PD 02-DEC-1998.
XX
PF 15-MAY-1997; 97JP-0140885.
XX
PR 15-MAY-1997; 97JP-0140885.
XX
PA (KURS) KURARAY CO LTD.
XX
DR WPI; 1999-076400/07.
XX
PT Material for medical treatment comprises new peptide - used for
PT covering injuries, promoting adhesion of bio-tissues, bone
PT reinforcing and nerve regeneration
XX
PS Claim 1; Page 12; 14pp; Japanese.
XX
CC The present invention describes a material for medical treatment which
CC comprises one or more peptides of the formula XADEGJLMProQY, or their
CC salts, immobilised on a substrate: where X = H, CH3CO or CH3COLys;
CC A = Ser or Thr; D = Ile, Val or Leu; E = Lys or Arg; G = Ile, Val or
CC Leu; J = Gly or Ala; L = Ile, Val or Leu; M = Gly or Ala; Q = Gly, Ala
CC or Gly-Lys-Lys-Gly; Y = OH or NH2. Also described is an agent for cell
CC growth promotion and/or cell adhesion promotion containing the above
CC peptide or its salt as the active component. The peptide and its salt
CC can be used for covering injuries, promoting adhesion of biotissues,
CC bone reinforcing and nerve regeneration. The present sequence represents
CC a specifically claimed peptide of the present invention.
XX
SQ Sequence 23 AA;

Query Match 100.0%; Score 69; DB 20; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db         12 DACEGDSGGPFV 23
```

RESULT 3
AAB12893
ID AAB12893 standard; peptide; 23 AA.
XX
AC AAB12893;
XX
DT 02-NOV-2000 (first entry)
XX
DE Nerve tissue regenerative peptide SEQ ID #8.
XX
KW Nerve regeneration; nerve cell proliferation; axon extension; treatment;
KW central nervous system disorder; peripheral nervous system disorder;
KW spinal disorder; head injury; cerebrovascular disorder.
XX
OS Synthetic.
XX
PN JP2000143531-A.
XX
PD 23-MAY-2000.
XX
PF 11-AUG-1999; 99JP-0227108.
XX
PR 09-SEP-1998; 98JP-0270498.
XX
PA (KURS) KURARAY CO LTD.
PA (NISH/) NISHIMURA Y.
PA (SUZU/) SUZUKI Y.
PA (TANI/) TANIHARA M.
XX
DR WPI; 2000-415772/36.
XX
PT New nerve regeneration material -
XX
PS Claim 2; Page 5; 17pp; Japanese.
XX
CC This invention relates to a new nerve regenerative material which
CC contains a peptide immobilised to a base which consists of a
CC polysaccharide gel such as alginic acid. Sequences AAB12886-B12899
CC represent examples of the peptides used in the nerve regeneration
CC material. The peptide containing material causes nerve cell
CC proliferation and also causes axonal extension. The material can be used
CC for the treatment of central or peripheral nervous system disorders,
CC spinal disorders, head injury or cerebrovascular disorders.
XX
SQ Sequence 23 AA;

```
  Query Match           100.0%;  Score 69;  DB 21;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 0.0051;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db         12 DACEGDSGGPFV 23
```

RESULT 4
AAB70363
ID AAB70363 standard; peptide; 23 AA.
XX
AC AAB70363;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human thrombin receptor binding domain peptide SEQ ID NO:8.
XX
KW Neutrophil cell chemotactic; wound healing; inflammation; vulnerary;
KW antiinflammatory.
XX
OS Homo sapiens.
XX
PN US6184342-B1.
XX
PD 06-FEB-2001.
XX
PF 28-OCT-1994; 94US-0330594.
XX
PR 28-OCT-1994; 94US-0330594.
XX
PA (CHRY-) CHRYSALIS BIOTECHNOLOGY INC.
XX
PI Carney DH, Ramakrishnan S;
XX
DR WPI; 2001-202003/20.
XX
PT New synthetic neutrophil cell chemotactic peptides, useful for
PT generating antibodies for modulating neutrophil chemotaxis in immune
PT response and wound healing -
XX
PS Example 2; Column 6; 15pp; English.
XX
CC The present invention describes a synthetic peptide (I) which is a
CC neutrophil cell chemotactic agent. (I) has vulnerary and
CC antiinflammatory activities. (I) is useful as a potent neutrophil cell
CC chemotactic agent and for generating antibodies against the peptides,
CC which are useful for modulating neutrophil recruitment to a wound site
CC for enhancing or inhibiting inflammation and early effects of wound
CC healing. Neutrophil response to (I) is specific, since monocytes and
CC fibroblasts do not show any expression of the receptor to which (I)
CC binds. The present sequence represents a human thrombin receptor binding
CC domain peptide which is used in an example from the present invention.
XX
SQ Sequence 23 AA;

```
  Query Match           100.0%;  Score 69;  DB 22;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 0.0051;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
```

```
Db         12 DACEGDSGGPFV 23

RESULT 5
AAE22563
ID   AAE22563 standard; peptide; 23 AA.
XX
AC   AAE22563;
XX
DT   26-JUL-2002  (first entry)
XX
DE   Human thrombin high affinity receptor binding domain.
XX
KW   Human; proteolytically activated receptor for thrombin; neutrophil;
KW   chemotactic agent; PART; inflammation; wound healing; chemotaxis;
KW   immune response; vulnerary; thrombin; receptor binding domain.
XX
OS   Homo sapiens.
XX
PN   US2002032314-A1.
XX
PD   14-MAR-2002.
XX
PF   05-FEB-2001; 2001US-0777328.
XX
PR   28-OCT-1994;   94US-0330594.
XX
PA   (CHRY-) CHRYSALIS BIOTECHNOLOGY INC.
XX
PI   Carney DH,  Ramakrishnan S;
XX
DR   WPI; 2002-371207/40.
XX
PT   New synthetic peptide neutrophil cell chemotactic agents, useful for
PT   stimulating or modulating neutrophil cell chemotactic migration,
PT   particularly for modulating neutrophil recruitment during immune
PT   response or in wound healing  -
XX
PS   Example 2; Page 3; 15pp; English.
XX
CC   The present invention relates to novel synthetic peptides and antibodies
CC   which are potent chemotactic agents for neutrophils. The peptides of the
CC   invention mimic the activity and role of the cleavage fragment of the
CC   proteolytically activated receptor for thrombin (PART). They are useful
CC   for stimulating or modulating neutrophil cell chemotactic migration or
CC   for generating an antibody. In particular, the peptides of the invention
CC   are useful for modulating neutrophil recruitment to a wound site for
CC   enhancing or inhibiting inflammation and early effects in wound healing.
CC   They are also useful for modulated neutrophil chemotaxis in immune
CC   response. The present sequence is high affinity receptor binding
CC   domain of human thrombin. This peptide is used in the exemplification
CC   of the invention.
XX
SQ   Sequence   23 AA;

  Query Match           100.0%;  Score 69;  DB 23;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 0.0051;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db         12 DACEGDSGGPFV 23


RESULT 6
AAE20159
ID   AAE20159 standard; peptide; 23 AA.
XX
AC   AAE20159;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human thrombin peptide derivative #2.
XX
KW   Cartilage growth; cartilage repair; arthritic joint; traumatic injury;
KW   non-proteolytically activated thrombin receptor; NPAR; chondrocyte;
KW   therapy; implantation; thrombin peptide; human.
XX
OS   Homo sapiens.
XX
PN   WO200207748-A2.
XX
PD   31-JAN-2002.
XX
PF   19-JUL-2001; 2001WO-US22668.
XX
PR   20-JUL-2000; 2000US-219800P.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Carney DH,  Crowther RS,  Stiernberg J,  Bergmann J;
XX
DR   WPI; 2002-268953/31.
XX
PT   Stimulating growth and repair of cartilage, useful for treating e.g.
PT   arthritis, by local administration of an agonist of non-proteolytically
PT   activated thrombin receptor  -
XX
PS   Claim 12; Page 25; 28pp; English.
XX
CC   The invention relates to a method of stimulating growth and repair of
CC   cartilage. The method involves administering to the site, an agonist
CC   of non-proteolytically activated thrombin receptor (NPAR). The method
CC   is used in human or veterinary medicine for the treatment of arthritic
CC   joints and damage/loss of cartilage caused by traumatic injury. Also
CC   chondrocytes may be cultured in presence of NPAR agonist to provide
CC   cells for implantation at sites requiring growth/repair of cartilage.
CC   The present sequence is human thrombin peptide derivative which serves
CC   as a NPAR agonist.
XX
SQ   Sequence   23 AA;

  Query Match           100.0%;  Score 69;  DB 23;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 0.0051;
```

```
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
  \.          ||||||||||||
Db         12 DACEGDSGGPFV 23


RESULT 7
AAU78376
ID   AAU78376 standard; Peptide; 23 AA.
XX
AC   AAU78376;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Thrombin peptide derivative TP508.
XX
KW   Thrombin; osteopathic; receptor; agonist; bone growth stimulation;
KW   osteoinduction; farm animal; companion animal; laboratory animal;
KW   bone graft; segmental bone gap; bone void; non-union fracture.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 3
FT                   /label= Unknown
XX
PN   WO200205836-A2.
XX
PD   24-JAN-2002.
XX
PF   18-JUL-2001; 2001WO-US22641.
XX
PR   19-JUL-2000; 2000US-219300P.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Carney DH,  Crowther RS,  Simmons DJ,  Yang J,  Redin WR;
XX
DR   WPI; 2002-303796/34.
XX
PT   Stimulating bone growth at a site in a subject in need of
PT   osteoinduction, such as a site of bone graft, segmental bone gap, bone
PT   void or non-union structure, by administering agonist of activated
PT   thrombin receptor  -
XX
PS   Claim 11; Page 22; 27pp; English.
XX
CC   The invention describes a method of stimulating bone growth at a site
CC   in a subject in need of osteoinduction. The method involves administering
CC   an agonist to stimulate bone growth at a site in a subject (e.g. a farm
CC   animal, companion animal or laboratory animal), in need of
CC   osteoinduction, such as the site in need of a bone graft in a subject, a
CC   segmental bone gap, a bone void or a non-union fracture. This sequence
CC   represents a thrombin peptide derivative obtained from a serine esterase
CC   that can stimulate or activate the non-proteolytically activated thrombin
CC   receptor.
XX
SQ   Sequence   23 AA;

  Query Match             100.0%;  Score 69;  DB 23;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 0.0051;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db         12 DACEGDSGGPFV 23


RESULT 8
AAM50858
ID   AAM50858 standard; Peptide; 23 AA.
XX
AC   AAM50858;
XX
DT   01-MAY-2002  (first entry)
XX
DE   Thrombin-derived peptide used to promote cardiac tissue repair.
XX
KW   Thrombin; revascularisation; vascular occlusion; tissue repair;
KW   vulnerary; vasotropic; cardiant; angiogenesis; restenosis;
KW   therapy; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         10..13
FT                   /note= "thrombin receptor binding domain"
FT   Peptide         12..23
FT                   /note= "serine esterase conserved sequence"
XX
PN   WO200204008-A2.
XX
PD   17-JAN-2002.
XX
PF   12-JUL-2001; 2001WO-US21944.
XX
PR   12-JUL-2000; 2000US-217583P.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Carney DH;
XX
DR   WPI; 2002-179665/23.
XX
PT   Promoting cardiac tissue repair, stimulating revascularisation,
PT   stimulating vascular endothelial cell proliferation, and inhibiting
PT   vascular occlusion by using angiogenic thrombin derivative peptide  -
XX
PS   Claim 4; Page 19; 24pp; English.
XX
CC   The present peptide comprises a thrombin-derived peptide, TP508,
CC   that includes a thrombin receptor binding domain sequence (see also
CC   AAM50856) and a serine esterase conserved sequence (see also
```

CC AAM50857). The peptide is used in a claimed method for promoting
CC cardiac tissue repair. It is administered during or following
CC cardiac surgery by injection into cardiac tissue, and may be
CC formulated as a sustained release formulation. The thrombin
CC derivative peptide is also used in claimed methods of stimulating
CC revascularisation, stimulating vascular endothelial cell
CC proliferation, inhibiting vascular occlusion, and inhibiting
CC restenosis following balloon angioplasty, in which case it may be
CC coated onto the catheter.
XX
SQ Sequence 23 AA;

Query Match 100.0%; Score 69; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
||||||||||||
Db 12 DACEGDSGGPFV 23

RESULT 9
ABP72755
ID ABP72755 standard; Peptide; 23 AA.
XX
AC ABP72755;
XX
DT 11-JUN-2003 (first entry)
XX
DE Antiulcer peptide derived from human thrombin.
XX
KW Antiulcer; human; thrombin.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal H or R3-C(O), where R3 is H or
FT a C1-C6 alkyl group"
FT Misc-difference 3
FT /note= "given as Try in the specification"
FT Modified-site 23
FT /note= "C-terminal OH or NR4R5, where R4 and R5 are
FT independently H, a C1-C6 alkyl group or,
FT taken together with the N atom to which they
FT are bonded, a non-aromatic heterocyclic
FT group"
FT Modified-site 1..23
FT /note= "0, 1, 2 or 3 amino acids at positions 1-9
FT and 14-23 differ from the given sequence
FT e.g. are conservative substitutions of the
FT amino acid at the corresponding position of
FT this sequence"
XX
PN WO2003013569-A2.
XX
PD 20-FEB-2003.
XX
PF 16-JAN-2002; 2002WO-US01151.
XX
PR 27-JUL-2001; 2001US-308198P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH;
XX
DR WPI; 2003-289898/28.
XX
PT Promoting healing of chronic dermal skin ulcer such as diabetic ulcer,
PT on a subject, by contacting the skin ulcer with an agonist of
PT non-proteolytically activated thrombin receptor -
XX
PS Claim 1; Page 14; 19pp; English.
XX
CC The present sequence is that of a human thrombin-derived peptide
CC based on prothrombin amino acid residues 508-530. The peptide acts
CC as an agonist of the non-proteolytically activated thrombin
CC receptor and has antiulcer activity. A claimed method of promoting
CC healing of a chronic dermal skin ulcer on a subject comprises
CC contacting the ulcer with an effective amount of this peptide, or an
CC N-terminal truncated fragment of it having at least 14 amino acids,
CC or a C-terminal truncated fragment of it having at least 18 amino
CC acids. Preferably, the peptide has -H at the N-terminus and -NH2 or
CC -OH at the C-terminus. An example is peptide TP508 (see ABP72757),
CC which was shown in an example from the invention to accelerate
CC the healing of chronic diabetic ulcers and to increase the
CC percentage of ulcer closure. The thrombin-derived peptides of the
CC invention can be used to treat a chronic dermal skin ulcer,
CC especially a diabetic ulcer, decubitus ulcer, venous stasis ulcer
CC or an arterial ulcer on a human, a companion animal, farm animal or
CC laboratory animal. They are inexpensive to produce and cause few,
CC if any, side effects.
XX
SQ Sequence 23 AA;

Query Match 100.0%; Score 69; DB 24; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
||||||||||||
Db 12 DACEGDSGGPFV 23

RESULT 10
ABP72757
ID ABP72757 standard; Peptide; 23 AA.
XX
AC ABP72757;
XX
DT 11-JUN-2003 (first entry)
XX
DE Antiulcer peptide TP508 derived from human thrombin.

XX
KW Antiulcer; human; thrombin.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Misc-difference 3
FT   /note= "given as Try in the specification"
FT Modified-site   23
FT   /note= "C-terminal amide"
XX
PN WO2003013569-A2.
XX
PD 20-FEB-2003.
XX
PF 16-JAN-2002; 2002WO-US01151.
XX
PR 27-JUL-2001; 2001US-308198P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH;
XX
DR WPI; 2003-289898/28.
XX
PT Promoting healing of chronic dermal skin ulcer such as diabetic ulcer,
PT on a subject, by contacting the skin ulcer with an agonist of
PT non-proteolytically activated thrombin receptor -
XX
PS Claim 15; Page 16; 19pp; English.
XX
CC The present sequence is that of a preferred human thrombin-derived
CC peptide of the invention that is based on prothrombin amino acid
CC residues 508-530. It is denoted TP508. The peptide acts as an
CC agonist of the non-proteolytically activated thrombin receptor and
CC has antiulcer activity. In an example from the invention, TP508
CC was shown to accelerate the healing of chronic diabetic ulcers and
CC to increase the percentage of ulcer closure. The antiulcer
CC peptides of the invention can be used to treat a chronic dermal
CC skin ulcer, especially a diabetic ulcer, decubitus ulcer, venous
CC stasis ulcer or an arterial ulcer on a human, a companion animal,
CC farm animal or laboratory animal. The peptides are inexpensive to
CC produce and cause few, if any, side effects.
XX
SQ Sequence 23 AA;

  Query Match   100.0%; Score 69; DB 24; Length 23;
  Best Local Similarity 100.0%; Pred. No. 0.0051;
  Matches 12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 DACEGDSGGPFV 12
              ||||||||||||
Db   12 DACEGDSGGPFV 23

RESULT 11

ABP72760
ID ABP72760 standard; Peptide; 23 AA.
XX
AC ABP72760;
XX
DT 11-JUN-2003 (first entry)
XX
DE Human thrombin peptide fragment.
XX
KW Antiulcer; human; thrombin.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Misc-difference 3
FT   /note= "given as Try in the specification"
XX
PN WO2003013569-A2.
XX
PD 20-FEB-2003.
XX
PF 16-JAN-2002; 2002WO-US01151.
XX
PR 27-JUL-2001; 2001US-308198P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH;
XX
DR WPI; 2003-289898/28.
XX
PT Promoting healing of chronic dermal skin ulcer such as diabetic ulcer,
PT on a subject, by contacting the skin ulcer with an agonist of
PT non-proteolytically activated thrombin receptor -
XX
PS Disclosure; Page 3; 19pp; English.
XX
CC The present sequence is that of a human thrombin-derived peptide
CC comprising prothrombin amino acid residues 508-530. The invention
CC provides peptides based on this sequence (see ABP72755-59) that act
CC as agonists of the non-proteolytically activated thrombin receptor
CC and which have antiulcer activity. One of these thrombin-derived
CC peptides (see ABP72756) was shown to accelerate the healing of
CC chronic diabetic ulcers and to increase the percentage of ulcer
CC closure. The peptides of the invention can be used to treat a
CC chronic dermal skin ulcer, especially a diabetic ulcer, decubitus
CC ulcer, venous stasis ulcer or an arterial ulcer on a human, a
CC companion animal, farm animal or laboratory animal. They are
CC inexpensive to produce and cause few, if any, side effects.
XX
SQ Sequence 23 AA;

  Query Match   100.0%; Score 69; DB 24; Length 23;
  Best Local Similarity 100.0%; Pred. No. 0.0051;
  Matches 12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 DACEGDSGGPFV 12

```
                  ||||||||||||
Db          12 DACEGDSGGPFV 23


RESULT 12
ABP72758
ID   ABP72758 standard; Peptide; 33 AA.
XX
AC   ABP72758;
XX
DT   11-JUN-2003  (first entry)
XX
DE   Antiulcer peptide derived from human thrombin.
XX
KW   Antiulcer; human; thrombin.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-terminal H or R3-C(O), where R3 is H or
FT                          a C1-C6 alkyl group"
FT   Misc-difference 8
FT                   /note= "given as Try in the specification"
FT   Modified-site   33
FT                   /note= "C-terminal OH or NR4R5, where R4 and R5 are
FT                          independently H, a C1-C6 alkyl group or,
FT                          taken together with the N atom to which they
FT                          are bonded, a non-aromatic heterocyclic
FT                          group"
FT   Modified-site   1..33
FT                   /note= "0, 1, 2 or 3 amino acids at positions 1-14
FT                          and 19-33 differ from the given sequence
FT                          e.g. are conservative substitutions of the
FT                          amino acid at the corresponding position of
FT                          this sequence"
XX
PN   WO2003013569-A2.
XX
PD   20-FEB-2003.
XX
PF   16-JAN-2002; 2002WO-US01151.
XX
PR   27-JUL-2001; 2001US-308198P.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Carney DH;
XX
DR   WPI; 2003-289898/28.
XX
PT   Promoting healing of chronic dermal skin ulcer such as diabetic ulcer,
PT   on a subject, by contacting the skin ulcer with an agonist of
PT   non-proteolytically activated thrombin receptor -
XX
PS   Claim 17; Page 16; 19pp; English.
XX
CC   The present sequence is that of a human thrombin-derived peptide
CC   that acts as an agonist of the non-proteolytically activated thrombin
CC   receptor.  It has antiulcer activity.  A claimed method of promoting
CC   healing of a chronic dermal skin ulcer on a subject comprises
CC   contacting the ulcer with an effective amount of this peptide, or an
CC   N-terminal truncated fragment of it having at least 14 amino acids,
CC   or a C-terminal truncated fragment of it having at least 18 amino
CC   acids.  Preferably, the peptide has -H at the N-terminus and -NH2 or
CC   -OH at the C-terminus.  The thrombin-derived peptides of the
CC   invention accelerate the healing of chronic diabetic ulcers and
CC   increase the percentage of ulcer closure.  They can be used to
CC   treat a chronic dermal skin ulcer, especially a diabetic ulcer,
CC   decubitus ulcer, venous stasis ulcer or an arterial ulcer on a
CC   human, a companion animal, farm animal or laboratory animal.  The
CC   peptides are inexpensive to produce and cause few, if any, side
CC   effects.
XX
SQ   Sequence   33 AA;

  Query Match             100.0%;  Score 69;  DB 24;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.0071;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db         17 DACEGDSGGPFV 28


RESULT 13
AAW99113
ID   AAW99113 standard; protein; 111 AA.
XX
AC   AAW99113;
XX
DT   14-MAY-1999  (first entry)
XX
DE   Bovine zeta 2 prethrombin 2.
XX
KW   Prothrombin; exosite assay; anticoagulant; blood clot; thrombin;
KW   cardiovascular disease; stroke; haematological disorder.
XX
OS   Bos sp.
XX
PN   WO9855130-A1.
XX
PD   10-DEC-1998.
XX
PF   28-MAY-1998;   98WO-US10840.
XX
PR   08-APR-1998;   98US-0081030.
PR   06-JUN-1997;   97US-0048864.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Krishnaswamy S;
XX
```

DR WPI; 1999-070237/06.
XX
PT Exosite assay for agents that inhibit catalytic cleavage of
PT prothrombin - at sites away from the active site of prothrombinase,
PT also new inhibitors, potentially useful as anticoagulants
XX
PS Disclosure; Page 42-43; 61pp; English.
XX
CC An exosite assay has been developed for inhibition of the catalytic
CC cleavage of prothrombin (PTh) to thrombin (Th) by prothrombinase (I), at
CC a site remote from the catalytic site of (I) comprises: (a) preparing a
CC solution containing 0.05-20 mu M substrate (S), that includes a protease
CC cleavage site and exosite-binding determinant; 0.05-200 nM factor Va;
CC 30-500 micro M phospholipids (PL); test inhibitor (A) in buffer of pH
CC 7-9, containing 1-10 mM calcium ions but no calcium-chelating agent;
CC (b) initiating catalytic cleavage of (S) by adding an aliquot of factor
CC Xa (to final concentration 0.05-200 nM) so that there is an excess of Va
CC over Xa, forming a S/(I) complex; (c) withdrawing aliquots of the
CC reaction mixture, quenching them; and (d) assaying for concentration of
CC Th. Alternatively, in the initial solution S is replaced by the same
CC concentration of Xa (less than the amount of Va), and reaction is started
CC by adding S. Also described in the present invention are inhibitors (A')
CC having IC50 less than 1 mu M identified by this assay. (A') are
CC potentially useful as a new class of anticoagulants for treatment of
CC cardiovascular disease, stroke and haematological disorders. The method
CC is based on the finding that exosite interactions are critical for
CC substrate specificity in catalytic formation of Th. The present sequence
CC represents bovine zeta 2 prethrombin 2.
XX
SQ Sequence 111 AA;

Query Match 100.0%; Score 69; DB 20; Length 111;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
||||||||||||
Db 51 DACEGDSGGPFV 62

RESULT 14
AAW99115
ID AAW99115 standard; protein; 116 AA.
XX
AC AAW99115;
XX
DT 14-MAY-1999 (first entry)
XX
DE Human zeta 2 prethrombin 2.
XX
KW Prothrombin; exosite assay; anticoagulant; blood clot; thrombin;
KW cardiovascular disease; stroke; haematological disorder.
XX
OS Homo sapiens.
XX
PN WO9855130-A1.
XX
PD 10-DEC-1998.
XX
PF 28-MAY-1998; 98WO-US10840.
XX
PR 08-APR-1998; 98US-0081030.
PR 06-JUN-1997; 97US-0048864.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Krishnaswamy S;
XX
DR WPI; 1999-070237/06.
XX
PT Exosite assay for agents that inhibit catalytic cleavage of
PT prothrombin - at sites away from the active site of prothrombinase,
PT also new inhibitors, potentially useful as anticoagulants
XX
PS Disclosure; Page 44-45; 61pp; English.
XX
CC An exosite assay has been developed for inhibition of the catalytic
CC cleavage of prothrombin (PTh) to thrombin (Th) by prothrombinase (I), at
CC a site remote from the catalytic site of (I) comprises: (a) preparing a
CC solution containing 0.05-20 mu M substrate (S), that includes a protease
CC cleavage site and exosite-binding determinant; 0.05-200 nM factor Va;
CC 30-500 micro M phospholipids (PL); test inhibitor (A) in buffer of pH
CC 7-9, containing 1-10 mM calcium ions but no calcium-chelating agent;
CC (b) initiating catalytic cleavage of (S) by adding an aliquot of factor
CC Xa (to final concentration 0.05-200 nM) so that there is an excess of Va
CC over Xa, forming a S/(I) complex; (c) withdrawing aliquots of the
CC reaction mixture, quenching them; and (d) assaying for concentration of
CC Th. Alternatively, in the initial solution S is replaced by the same
CC concentration of Xa (less than the amount of Va), and reaction is started
CC by adding S. Also described in the present invention are inhibitors (A')
CC having IC50 less than 1 mu M identified by this assay. (A') are
CC potentially useful as a new class of anticoagulants for treatment of
CC cardiovascular disease, stroke and haematological disorders. The method
CC is based on the finding that exosite interactions are critical for
CC substrate specificity in catalytic formation of Th. The present sequence
CC represents human zeta 2 prethrombin 2.
XX
SQ Sequence 116 AA;

Query Match 100.0%; Score 69; DB 20; Length 116;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
||||||||||||
Db 56 DACEGDSGGPFV 67

RESULT 15
AAW11545
ID AAW11545 standard; Protein; 259 AA.
XX
AC AAW11545;
XX

DT 01-OCT-1997 (first entry)
XX
DE Human thrombin Asn99 mutant.
XX
KW Prothrombin; mutant; mutein; platelet aggregation; blood clotting;
KW coagulation; reduce; decrease; hirudin; heparin; anti-thrombin III;
KW antagonist; D99N.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Protein 1..259
FT /label= thrombin_Asn99
FT Misc-difference 99
FT /note= "Wild-type Asp residue has been replaced by
FT Asn"
XX
PN WO9641868-A2.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-AT00105.
XX
PR 13-JUN-1995; 95AT-0001006.
XX
PA (IMMO) IMMUNO AG.
XX
PI Eibl J, Falkner F, Fischer B, Mitterer A, Schlokat U;
XX
DR WPI; 1997-065455/06.
XX
PT Prothrombin mutants with reduced clotting activity - useful as
PT antagonists of thrombin inhibitors or for anticoagulant therapy
XX
PS Example 3; Page -; 73pp; German.
XX
CC Prothrombin mutants having one or more changes in amino acid sequence
CC compared with the natural protein and having 0-10% (preferably 0-0.25%)
CC of the activity of the natural protein are claimed, provided that the
CC changes in amino acid sequence do not affect the capacity of the
CC mutants to bind to specific ligands and receptors. The mutants have
CC greatly reduced clotting activity and are useful as antagonists of
CC thrombin inhibitors such as hirudin, heparin and anti-thrombin III.
CC The mutations may also result in changes to the in vivo half-life
CC of prothrombin. The half-life may be reduced to less than 10 minutes
CC or the mutant prothrombin may have an extended half-life of more than
CC 1 hour, making it useful as an anticoagulant and to inhibit side-
CC effects of anti-coagulant treatment. They are converted to inactive
CC thrombin and are able to compete with native, active thrombin for
CC binding to receptors. The present sequence represents the thrombin
CC mutant which is derived by trypsin cleavage of a specifically
CC claimed human prothrombin mutant in which Asp at position 419 is
CC changed to Asn. The thrombin Asn99 mutant was found to have only
CC 0.24% of the activity of wild-type thrombin on a chromogenic
CC substrate.
CC (Note: This sequence does not appear in the specification and has
CC been produced by modifying the wild-type sequence of human
CC prothrombin which appears in figure 1).
XX
SQ Sequence 259 AA;

Query Match 100.0%; Score 69; DB 18; Length 259;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        199 DACEGDSGGPFV 210
```

Search completed: February 11, 2004, 14:53:24
Job time : 25.9355 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:49:07 ; Search time 8.12903 Seconds
(without alignments)
141.963 Million cell updates/sec

Title: US-10-050-611-2
Perfect score: 69
Sequence: 1 DACEGDSGGPFV 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	100.0	234	2	F42696	thrombin (EC 3.4.2
2	69	100.0	235	2	D42696	thrombin (EC 3.4.2
3	69	100.0	235	2	E42696	thrombin (EC 3.4.2
4	69	100.0	236	2	C42696	thrombin (EC 3.4.2
5	69	100.0	236	2	I42696	thrombin (EC 3.4.2
6	69	100.0	239	2	G42696	thrombin (EC 3.4.2
7	69	100.0	617	2	S10511	thrombin (EC 3.4.2
8	69	100.0	618	2	A35827	thrombin (EC 3.4.2
9	69	100.0	622	1	TBHU	thrombin (EC 3.4.2
10	69	100.0	625	1	TBBO	thrombin (EC 3.4.2
11	66	95.7	417	1	S00845	hepsin (EC 3.4.21.
12	66	95.7	1524	2	T30337	polyprotein - Afri
13	63	91.3	235	2	H42696	thrombin (EC 3.4.2
14	63	91.3	456	1	KXBO	protein C (activat
15	63	91.3	461	1	KXHU	protein C (activat
16	60	87.0	254	2	S65465	trypsin-like prote
17	60	87.0	256	1	TRFF	trypsin-like prote
18	60	87.0	264	2	S32794	trypsin-like prote
19	60	87.0	267	2	S40006	trypsin (EC 3.4.21
20	60	87.0	271	2	S41308	serine proteinase
21	60	87.0	274	2	S35339	trypsin (EC 3.4.21
22	60	87.0	275	2	S40007	trypsin (EC 3.4.21
23	60	87.0	275	2	S40005	trypsin (EC 3.4.21
24	60	87.0	277	2	S35340	trypsin (EC 3.4.21
25	60	87.0	285	2	T35195	probable serine pr
26	60	87.0	394	2	JS0600	t-plasminogen acti
27	60	87.0	431	2	JS0599	t-plasminogen acti
28	60	87.0	461	1	S18994	protein C (activat
29	60	87.0	461	1	JX0210	protein C (activat
30	60	87.0	477	1	A34369	t-plasminogen acti
31	60	87.0	477	2	JS0597	t-plasminogen acti
32	60	87.0	477	2	JS0598	t-plasminogen acti
33	60	87.0	559	1	A35029	t-plasminogen acti
34	60	87.0	559	1	A29941	t-plasminogen acti
35	60	87.0	562	1	UKHUT	t-plasminogen acti
36	60	87.0	593	2	S45281	coagulation factor
37	60	87.0	603	2	S28941	coagulation factor
38	60	87.0	615	1	KFHU12	coagulation factor
39	59	85.5	161	2	I62744	coagulation factor
40	59	85.5	161	2	I48158	coagulation factor
41	59	85.5	265	2	T15451	hypothetical prote
42	59	85.5	275	2	I46712	factor IX - rabbit
43	59	85.5	282	2	I84621	coagulation factor
44	59	85.5	434	1	A35005	u-plasminogen acti
45	59	85.5	459	2	JQ0419	coagulation factor

ALIGNMENTS

RESULT 1
F42696
thrombin (EC 3.4.21.5) B chain - Cynops pyrogastor (fire-bellied newt) (fragment)
C;Species: Cynops pyrogastor (fire-bellied newt)
C;Date: 19-Mar-1997 #sequence_revision 19-Dec-1997 #text_change 17-Mar-1999
C;Accession: F42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Note: sequence not
A;Accession: F42696
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-234 <BAN>
A;Cross-references: GB:M81395
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology

C;Keywords: hydrolase; serine proteinase

  Query Match 100.0%; Score 69; DB 2; Length 234;
  Best Local Similarity 100.0%; Pred. No. 0.00051;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
      ||||||||||||
Db 174 DACEGDSGGPFV 185

RESULT 2
D42696
thrombin (EC 3.4.21.5) B chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 17-Mar-1999
C;Accession: D42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: D42696
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-235 <BAN>
A;Cross-references: GB:M81391
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-226/Domain: trypsin homology (fragment) <TRY>

  Query Match 100.0%; Score 69; DB 2; Length 235;
  Best Local Similarity 100.0%; Pred. No. 0.00052;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
      ||||||||||||
Db 175 DACEGDSGGPFV 186

RESULT 3
E42696
thrombin (EC 3.4.21.5) B chain - tokay (fragment)
C;Species: Gekko gecko (tokay)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 17-Mar-1999
C;Accession: E42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: E42696
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-235 <BAN>
A;Cross-references: GB:M81392
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-226/Domain: trypsin homology (fragment) <TRY>

  Query Match 100.0%; Score 69; DB 2; Length 235;
  Best Local Similarity 100.0%; Pred. No. 0.00052;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
      ||||||||||||
Db 175 DACEGDSGGPFV 186

RESULT 4
C42696
thrombin (EC 3.4.21.5) B chain - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 17-Mar-1999
C;Accession: C42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: C42696
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-236 <BAN>
A;Cross-references: GB:M81396
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-227/Domain: trypsin homology (fragment) <TRY>

  Query Match 100.0%; Score 69; DB 2; Length 236;
  Best Local Similarity 100.0%; Pred. No. 0.00052;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
      ||||||||||||
Db 176 DACEGDSGGPFV 187

RESULT 5
I42696
thrombin (EC 3.4.21.5) B chain - Pacific hagfish (fragment)
C;Species: Eptatretus stouti (Pacific hagfish)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 17-Mar-1999
C;Accession: I42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: I42696
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA

A;Residues: 1-236 <BAN>
A;Cross-references: GB:M81393
A;Note: nuclectide translation not given
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-226/Domain: trypsin homology (fragment) <TRY>

  Query Match   100.0%; Score 69; DB 2; Length 236;
  Best Local Similarity   100.0%; Pred. No. 0.00052;
  Matches   12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 DACEGDSGGPFV 12
              ||||||||||||
Db   175 DACEGDSGGPFV 186

RESULT 6
G42696
thrombin (EC 3.4.21.5) B chain - rainbow trout (fragment)
C;Species: Oncorhynchus mykiss (rainbow trout)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 22-Jun-1999
C;Accession: G42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification
and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: G42696
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-239 <BAN>
A;Cross-references: GB:M81398; NID:g213486; PIDN:AAA49433.1; PID:g213487
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-226/Domain: trypsin homology (fragment) <TRY>

  Query Match   100.0%; Score 69; DB 2; Length 239;
  Best Local Similarity   100.0%; Pred. No. 0.00052;
  Matches   12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 DACEGDSGGPFV 12
              ||||||||||||
Db   175 DACEGDSGGPFV 186

RESULT 7
S10511
thrombin (EC 3.4.21.5) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 03-May-2002
C;Accession: S10511; A60576; B42696
R;Dihanich, M.; Monard, D.
Nucleic Acids Res. 18, 4251, 1990
A;Title: cDNA sequence of rat prothrombin.
A;Reference number: S10511; MUID:90332426; PMID:2377469
A;Accession: S10511
A;Molecule type: mRNA
A;Residues: 1-617 <DIH>
A;Cross-references: EMBL:X52835; NID:g56969; PIDN:CAA37017.1; PID:g56970
R;Henrikson, K.P.; Jazin, E.E.; Greenwood, J.A.; Dickerman, H.W.
Endocrinology 126, 167-175, 1990
A;Title: Prothrombin levels are increased in the estrogen-treated immature rat
uterus.
A;Reference number: A60576; MUID:90091942; PMID:2293980
A;Accession: A60576
A;Molecule type: protein
A;Residues: 44-58 <HEN>
A;Note: the authors purified the proenzyme from the estrogen-stimulated maturing
rat uterus and demonstrated it to be prothrombin
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification
and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: B42696
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 383-617,'E' <BAN>
A;Cross-references: GB:M81397
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: blood coagulation; calcium binding; carboxyglutamic acid;
glycoprotein; hydrolase; kringle; serine proteinase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-43/Domain: propeptide #status predicted <PRO>
F;28-88/Domain: Gla domain homology <GLA>
F;44-617/Product: prothrombin #status experimental <PMAT>
F;109-187/Domain: kringle homology <KR1>
F;215-292/Domain: kringle homology <KR2>
F;360-609/Domain: trypsin homology <TRY>
F;50,51,58,60,63,64,69,70,73,76/Modified site: gamma-carboxyglutamic acid (Glu)
#status predicted
F;61-66,91-104,109-187,130-170,158-182,215-292,236-276,264-287,332-478,387-
403,532-546,560-590/Disulfide bonds: #status predicted
F;402,458,564/Active site: His, Asp, Ser #status predicted

  Query Match   100.0%; Score 69; DB 2; Length 617;
  Best Local Similarity   100.0%; Pred. No. 0.0013;
  Matches   12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 DACEGDSGGPFV 12
              ||||||||||||
Db   558 DACEGDSGGPFV 569

RESULT 8
A35827
thrombin (EC 3.4.21.5) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 14-Dec-1990 #sequence_revision 14-Dec-1990 #text_change 03-May-2002
C;Accession: A35827; A42696; S12081
R;Degen, S.J.F.; Schaefer, L.A.; Jamison, C.S.; Grant, S.G.; Fitzgibbon, J.J.;
Pai, J.A.; Chapman, V.M.; Elliott, R.W.
DNA Cell Biol. 9, 487-498, 1990

A;Title: Characterization of the cDNA coding for mouse prothrombin and localization of the gene on mouse chromosome 2.
A;Reference number: A35827; MUID:91025551; PMID:2222810
A;Accession: A35827
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-618 <DEG>
A;Cross-references: GB:X52308; NID:g53813; PIDN:CAA36548.1; PID:g53814
A;Experimental source: strain C57BL/6
A;Note: the data were obtained from females resulting from the cross of M. domesticus and M. spretus
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: A42696
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 384-618,'E' <BAN>
A;Cross-references: GB:M81394
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: blood coagulation; calcium binding; carboxyglutamic acid; glycoprotein; hydrolase; kringle; serine proteinase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-43/Domain: propeptide #status predicted <PRO>
F;28-88/Domain: Gla domain homology <GLA>
F;44-618/Product: prothrombin B #status predicted <MAT>
F;109-187/Domain: kringle homology <KR1>
F;215-293/Domain: kringle homology <KR2>
F;361-610/Domain: trypsin homology <TRY>
F;50,51,58,60,63,64,69,70,73,76/Modified site: gamma-carboxyglutamic acid (Glu) #status predicted
F;61-66,91-104,109-187,130-170,158-182,215-293,236-276,264-288,333-479,388-404,533-547,561-591/Disulfide bonds: #status predicted
F;403,459,565/Active site: His, Asp, Ser #status predicted

  Query Match 100.0%; Score 69; DB 2; Length 618;
  Best Local Similarity 100.0%; Pred. No. 0.0013;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DACEGDSGGPFV 12
    ||||||||||||
Db 559 DACEGDSGGPFV 570

RESULT 9
TBHU
thrombin (EC 3.4.21.5) precursor [validated] - human
N;Alternate names: coagulation factor II
N;Contains: prothrombin
C;Species: Homo sapiens (man)
C;Date: 30-Nov-1980 #sequence_revision 22-Jul-1994 #text_change 08-Dec-2000
C;Accession: A29351; A00914; B00914; A37549; A37550; I51952
R;Degen, S.J.F.; Davie, E.W.
Biochemistry 26, 6165-6177, 1987
A;Title: Nucleotide sequence of the gene for human prothrombin.
A;Reference number: A29351; MUID:88077877; PMID:2825773
A;Accession: A29351
A;Molecule type: DNA
A;Residues: 1-622 <DEG>
A;Cross-references: GB:M17262; GB:M33691; NID:g558069; PIDN:AAC63054.1; PID:g339641
R;Degen, S.J.F.; MacGillivray, R.T.A.; Davie, E.W.
Biochemistry 22, 2087-2097, 1983
A;Title: Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin.
A;Reference number: A00914; MUID:83231469; PMID:6305407
A;Accession: A00914
A;Molecule type: mRNA
A;Residues: 8-163,'N',165-622 <DE2>
A;Cross-references: GB:V00595; GB:J00307; NID:g37128; PIDN:CAA23842.1; PID:g1335344
A;Accession: B00914
A;Molecule type: DNA
A;Residues: 188-311 <DE3>
R;Walz, D.A.; Hewett-Emmett, D.; Seegers, W.H.
Proc. Natl. Acad. Sci. U.S.A. 74, 1969-1972, 1977
A;Reference number: A37549; MUID:77193964; PMID:266717
A;Accession: A37549
A;Molecule type: protein
A;Residues: 44-118,'N',120,'S',122-163,'I',165-175,'A',177-182,'T',184-193,'MV',196-308,'EE',309-314 <WAL>
R;Butkowski, R.J.; Elion, J.; Downing, M.R.; Mann, K.G.
J. Biol. Chem. 252, 4942-4957, 1977
A;Title: Primary structure of human prethrombin 2 and alpha-thrombin.
A;Reference number: A37550; MUID:77207112; PMID:873923
A;Accession: A37550
A;Molecule type: protein
A;Residues: 315-334,'N',336-348,'N',350-368,'N',370-397,'N',399-413,'N',415-484,'N',486-493,'G',495-503,'Y',505-508,'S',510,'V',512-513,'D',515-528,'AL',531,'Q',533-622 <BUT>
R;Rabiet, M.J.; Blashill, A.; Furie, B.; Furie, B.C.
J. Biol. Chem. 261, 13210-13215, 1986
A;Reference number: A37551; MUID:87008532; PMID:3759958
A;Contents: annotation; activation cleavages
R;MacGillivray, R.T.; Irwin, D.M.; Guinto, E.R.; Stone, J.C.
Ann. N. Y. Acad. Sci. 485, 73-79, 1986
A;Title: Recombinant genetic approaches to functional mapping of thrombin.
A;Reference number: I51952; MUID:87182874; PMID:3471151
A;Accession: I51952
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2,'RI',5-100 <RES>
A;Cross-references: GB:M33031; NID:g190723; PIDN:AAA60220.1; PID:g190724
C;Comment: Thrombin, which cleaves bonds after Arg and Lys, converts fibrinogen to fibrin and activates factors V, VIII, XIII, and, in complex with thrombomodulin, protein C.
C;Comment: Prothrombin is activated on the surface of a phospholipid membrane that binds the amino end of prothrombin and factors Va and Xa in calcium-dependent interactions. The activation peptide(s) can be removed either by factor Xa or thrombin; the cleavage into light and heavy chains is by factor Xa. It is not known whether one or two smaller activation peptides, with additional cleavage after 314-Arg, are released in natural blood clotting.

C;Comment: The cleavage after Arg-198, observed in vitro, does not occur in plasma.
C;Comment: The gamma-carboxyglutamyl residues bind calcium ions, result from the carboxylation of glutamyl residues by microsomal vitamin K-dependent carboxylase, and are necessary for calcium-dependent interaction with the negatively charged phospholipid membrane surface.
C;Comment: The prothrombin precursor is synthesized in the liver.
C;Genetics:
A;Gene: GDB:F2
A;Cross-references: GDB:119894; OMIM:176930
A;Map position: 11p11-11q12
A;Introns: 27/1; 80/3; 89/1; 106/1; 141/2; 187/1; 292/1; 335/1; 377/2; 433/2; 491/2; 552/1; 575/3
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: acute phase; blood coagulation; calcium binding; carboxyglutamic acid; duplication; glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-43/Domain: propeptide #status predicted <PRO>
F;28-87/Domain: Gla domain homology <GLA>
F;44-622/Product: prothrombin #status experimental <MAT>
F;44-327/Domain: activation peptide #status experimental <APT>
F;108-186/Domain: kringle homology <KR1>
F;213-291/Domain: kringle homology <KR2>
F;328-363/Product: thrombin light chain #status experimental <LCH>
F;364-622/Product: thrombin heavy chain #status experimental <HCH>
F;364-613/Domain: trypsin homology <TRY>
F;49,50,57,59,62,63,68,69,72,75/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental
F;60-65,90-103,108-186,129-169,157-181,213-291,234-274,262-286/Disulfide bonds: #status predicted
F;121,143/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;336-482,536-550,564-594/Disulfide bonds: #status predicted
F;391-407/Disulfide bonds: #status experimental
F;406,462/Active site: His, Asp #status predicted
F;416/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;568/Active site: Ser #status experimental

  Query Match 100.0%; Score 69; DB 1; Length 622;
  Best Local Similarity 100.0%; Pred. No. 0.0013;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        562 DACEGDSGGPFV 573
```

RESULT 10
TBBO
thrombin (EC 3.4.21.5) precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 14-Jul-1994 #text_change 18-Jun-1999
C;Accession: S02537; A00915; A37552; I46045; S67518
R;Irwin, D.M.; Robertson, K.A.; MacGillivray, R.T.A.
J. Mol. Biol. 200, 31-45, 1988
A;Title: Structure and evolution of the bovine prothrombin gene.
A;Reference number: S02537; MUID:88245190; PMID:3379642
A;Accession: S02537
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-625 <IRW>
R;MacGillivray, R.T.A.; Davie, E.W.
Biochemistry 23, 1626-1634, 1984
A;Title: Characterization of bovine prothrombin mRNA and its translation product.
A;Reference number: A00915; MUID:84203525; PMID:6326805
A;Accession: A00915
A;Molecule type: mRNA
A;Residues: 1-230,'H',232-625 <MAC>
A;Note: 600-Asn was also found
R;Magnusson, S.; Sottrup-Jensen, L.; Petersen, T.E.; Claeys, H.
in Boerhaave Symposium on Prothrombin and Related Coagulation Factors, Hemker, H.C., and Veltkamp, J.J., eds., pp.25-46, Leiden Univ. Press, Leiden, 1975
A;Reference number: A37552
A;Accession: A37552
A;Molecule type: protein
A;Residues: 44-287,'N',289-352,'E',354,'Q',356-548,'ND',551-599,'N',601-625 <MAG>
A;Note: the evidence for 231-Ser is strong
A;Note: disulfide bonds and carbohydrate binding sites were determined
R;Park, C.H.; Tulinsky, A.
Biochemistry 25, 3977-3982, 1986
A;Title: Three-dimensional structure of the kringle sequence: structure of prothrombin fragment 1.
A;Reference number: A37553; MUID:86296631; PMID:3741841
A;Contents: annotation; residues 44-317, X-ray crystallography, 2.8 angstroms
R;Irwin, D.M.; Ahern, K.G.; Pearson, G.D.; MacGillivray, R.T.A.
Biochemistry 24, 6854-6861, 1985
A;Title: Characterization of the bovine prothrombin gene.
A;Reference number: A37554; MUID:86077733; PMID:3000440
A;Contents: annotation; gene structure
R;MacGillivray, R.T.; Degen, S.J.; Chandra, T.; Woo, S.L.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 77, 5153-5157, 1980
A;Title: Cloning and analysis of a cDNA coding for bovine prothrombin.
A;Reference number: I46045; MUID:81054926; PMID:6254059
A;Accession: I46045
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 466-599,'N',601-625 <MA2>
A;Cross-references: EMBL:V00135; NID:g772; PIDN:CAA23451.1; PID:g808945
R;Pejler, G.; Karlstroem, A.R.; Berg, L.
Eur. J. Biochem. 227, 102-107, 1995
A;Title: Identification of the proteolytic thrombin fragments formed after cleavage with rat mast cell protease 1.
A;Reference number: S67518; MUID:95154277; PMID:7851376
A;Accession: S67518
A;Status: preliminary
A;Molecule type: protein
A;Residues: 318-325;333-338,'X',340;367-374;481-484,'X',486-488;515-522 <PEJ>
C;Comment: Thrombin, which cleaves bonds after Arg and Lys, converts fibrinogen to fibrin and activates factors V, VII, VIII, XIII, and, in complex with thrombomodulin, protein C.
C;Comment: Prothrombin is activated on the surface of a phospholipid membrane that binds the amino end of prothrombin and factors Va and Xa in calcium-

dependent interactions; factor Xa removes the activation peptide and cleaves the remaining part into light and heavy chains. The activation process starts slowly because factor V itself has to be activated by the initial, small amounts of thrombin.
C;Comment: Thrombin can cleave the amino-terminal activation peptide 1 from prothrombin, prior to its activation by factor Xa.
C;Comment: The gamma-carboxyglutamyl residues bind calcium ions, result from the carboxylation of glutamyl residues by microsomal vitamin K-dependent carboxylase, and are necessary for calcium-dependent interaction with the negatively charged phospholipid membrane surface.
C;Comment: The prothrombin precursor is synthesized in the liver.
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: blood coagulation; calcium binding; carboxyglutamic acid; duplication; glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-43/Domain: propeptide #status predicted <PRO>
F;28-88/Domain: Gla domain homology <GLA>
F;44-625/Product: prothrombin #status experimental <MPT>
F;44-199/Domain: activation peptide 1 #status experimental <FR1>
F;109-187/Domain: kringle homology <KR1>
F;200-317/Domain: activation peptide 2 #status experimental <FR2>
F;214-292/Domain: kringle homology <KR2>
F;318-366/Product: thrombin light chain #status experimental <LCH>
F;367-625/Product: thrombin heavy chain #status experimental <HCH>
F;367-616/Domain: trypsin homology <TRY>
F;50,51,58,60,63,64,69,70,73,76/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental
F;61-66,91-104,109-187,130-170,158-182,214-292,235-275,263-287,339-485,394-410,539-553,567-597/Disulfide bonds: #status experimental
F;120,144,419/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;409,465,571/Active site: His, Asp, Ser #status experimental

  Query Match 100.0%; Score 69; DB 1; Length 625;
  Best Local Similarity 100.0%; Pred. No. 0.0013;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        565 DACEGDSGGPFV 576
```

RESULT 11
S00845
hepsin (EC 3.4.21.-) - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C;Accession: S00845
R;Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A;Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells.
A;Reference number: S00845; MUID:88209431; PMID:2835076
A;Accession: S00845
A;Molecule type: mRNA
A;Residues: 1-417 <LEY>
A;Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
C;Genetics:
A;Gene: GDB:HPN; TMPRSS1; hepsin
A;Cross-references: GDB:135685; OMIM:142440
A;Map position: 19q11-19q13.2
C;Superfamily: hepsin; trypsin homology
C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F;23-45/Domain: transmembrane #status predicted <TMN>
F;163-400/Domain: trypsin homology <TRY>
F;188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F;203,257,353/Active site: His, Asp, Ser #status predicted

  Query Match 95.7%; Score 66; DB 1; Length 417;
  Best Local Similarity 91.7%; Pred. No. 0.0028;
  Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DACEGDSGGPFV 12
              |||:||||||||
Db        347 DACQGDSGGPFV 358
```

RESULT 12
T30337
polyprotein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
C;Accession: T30337
R;Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A;Description: cDNA cloning of ovochymase, a chymotrypsin-like protease released from Xenopus laevis eggs at fertilization.
A;Reference number: Z20829
A;Accession: T30337
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1524 <YAN>
A;Cross-references: EMBL:U81290; NID:g2981640; PID:g2981641; PIDN:AAC24717.1
C;Superfamily: tryosin related polyprotein; trypsin homology

  Query Match 95.7%; Score 66; DB 2; Length 1524;
  Best Local Similarity 91.7%; Pred. No. 0.0093;
  Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DACEGDSGGPFV 12
              |||:||||||||
Db        241 DACQGDSGGPFV 252
```

Search completed: February 11, 2004, 14:56:56
Job time : 8.12903 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:36:52 ; Search time 5.03226 Seconds
(without alignments)
112.141 Million cell updates/sec

Title: US-10-050-611-2
Perfect score: 69
Sequence: 1 DACEGDSGGPFV 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	100.0	617	1	THRB_RAT	P18292 rattus norv
2	69	100.0	618	1	THRB_MOUSE	P19221 mus musculu
3	69	100.0	622	1	THRB_HUMAN	P00734 homo sapien
4	69	100.0	625	1	THRB_BOVIN	P00735 bos taurus
5	66	95.7	417	1	HEPS_HUMAN	P05981 homo sapien
6	66	95.7	436	1	HEPS_MOUSE	O35453 mus musculu
7	63	91.3	157	1	PRTC_CANFA	Q28278 canis famil
8	63	91.3	157	1	PRTC_CAPHI	Q28315 capra hircu
9	63	91.3	157	1	PRTC_FELCA	Q28412 felis silve
10	63	91.3	157	1	PRTC_HORSE	Q28380 equus cabal
11	63	91.3	161	1	PRTC_MACMU	Q28506 macaca mula
12	63	91.3	456	1	PRTC_BOVIN	P00745 bos taurus
13	63	91.3	459	1	PRTC_PIG	Q9glp2 sus scrofa
14	63	91.3	461	1	PRTC_HUMAN	P04070 homo sapien
15	60	87.0	248	1	KLKC_HUMAN	Q9ukr0 homo sapien
16	60	87.0	253	1	TRYB_DROER	P54625 drosophila
17	60	87.0	253	1	TRYD_DROER	P54626 drosophila
18	60	87.0	253	1	TRYD_DROME	P42276 drosophila
19	60	87.0	253	1	TRYG_DROME	P42277 drosophila
20	60	87.0	254	1	TRYP_SARBU	P51588 sarcophaga
21	60	87.0	256	1	HYPB_HYPLI	P35588 hypoderma l
22	60	87.0	256	1	TRYA_DROER	P54624 drosophila
23	60	87.0	256	1	TRYA_DROME	P04814 drosophila
24	60	87.0	256	1	TRYE_DROER	P54627 drosophila
25	60	87.0	256	1	TRYE_DROME	P35005 drosophila
26	60	87.0	258	1	TRYU_DROER	P54629 drosophila
27	60	87.0	262	1	TRYU_DROME	P42279 drosophila
28	60	87.0	264	1	VDP_BOMMO	Q07943 bombyx mori
29	60	87.0	267	1	TRY7_ANOGA	P35041 anopheles g
30	60	87.0	274	1	TRY1_ANOGA	P35035 anopheles g
31	60	87.0	275	1	TRY3_ANOGA	P35037 anopheles g
32	60	87.0	275	1	TRY4_ANOGA	P35038 anopheles g
33	60	87.0	277	1	KLKD_HUMAN	Q9ukr3 homo sapien
34	60	87.0	277	1	TRY2_ANOGA	P35036 anopheles g
35	60	87.0	281	1	TRYZ_DROER	P54630 drosophila
36	60	87.0	394	1	URTG_DESRO	P49150 desmodus ro
37	60	87.0	418	1	HATT_HUMAN	O60235 homo sapien
38	60	87.0	422	1	DES1_HUMAN	Q9ul52 homo sapien
39	60	87.0	431	1	URTB_DESRO	P98121 desmodus ro
40	60	87.0	455	1	TMS5_MOUSE	Q9er04 mus musculu
41	60	87.0	457	1	TMS5_HUMAN	Q9h3s3 homo sapien
42	60	87.0	458	1	PRTC_RABIT	Q28661 oryctolagus
43	60	87.0	461	1	PRTC_MOUSE	P33587 mus musculu
44	60	87.0	461	1	PRTC_RAT	P31394 rattus norv
45	60	87.0	477	1	URT1_DESRO	P98119 desmodus ro

ALIGNMENTS

RESULT 1
THRB_RAT
ID THRB_RAT STANDARD; PRT; 617 AA.
AC P18292;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Prothrombin precursor (EC 3.4.21.5).
GN F2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=90332426; PubMed=2377469;
RA Dihanich M., Monard D.;
RT "cDNA sequence of rat prothrombin.";
RL Nucleic Acids Res. 18:4251-4251(1990).
RN [2]
RP SEQUENCE OF 383-617 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92212913; PubMed=1557383;

```
RA   Banfield D.K., Macgillivray R.T.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Gly; activates
CC       fibrinogen to fibrin and releases fibrinopeptide A and B.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION
CC       OF PROTHROMBIN TO THROMBIN.
CC   -!- MISCELLANEOUS: PROTHROMBIN IS ACTIVATED ON THE SURFACE OF A
CC       PHOSPHOLIPID MEMBRANE THAT BINDS THE AMINO END OF PROTHROMBIN &
CC       FACTORS VA & XA IN CA-DEPENDENT INTERACTIONS; FACTOR XA REMOVES
CC       THE ACTIVATION PEPTIDE & CLEAVES THE REMAINING PART INTO LIGHT &
CC       HEAVY CHAINS. THE ACTIVATION PROCESS STARTS SLOWLY BECAUSE FACTOR
CC       V ITSELF HAS TO BE ACTIVATED BY THE INITIAL, SMALL AMOUNTS OF
CC       THROMBIN.
CC   -!- MISCELLANEOUS: THROMBIN CAN ITSELF CLEAVE THE AMINO TERMINAL
CC       FRAGMENT (FRAGMENT 1) OF THE PROTHROMBIN, PRIOR TO ITS ACTIVATION
CC       BY FACTOR XA.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   -!- SIMILARITY: Contains 2 kringle domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52835; CAA37017.1; -.
DR   EMBL; M81397; AAA42240.1; -.
DR   PIR; S10511; S10511.
DR   HSSP; P00734; 1UVS.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00051; kringle; 2.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00001; GLABLOOD.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   ProDom; PD000395; Kringle; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00130; KR; 2.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00021; KRINGLE_1; 2.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Blood coagulation; Plasma; Calcium-binding; Glycoprotein; Repeat;
KW   Vitamin K; Zymogen; Gamma-carboxyglutamic acid; Acute phase; Liver;
KW   Hydrolase; Serine protease; Kringle; Signal.
FT   SIGNAL       1    24       POTENTIAL.
FT   PROPEP      25    43
FT   CHAIN       44   617       PROTHROMBIN.
FT   PEPTIDE     44   200       ACTIVATION PEPTIDE (FRAGMENT 1).
FT   PEPTIDE    201   323       ACTIVATION PEPTIDE (FRAGMENT 2).
FT   CHAIN      324   359       THROMBIN LIGHT CHAIN (A).
FT   CHAIN      360   617       THROMBIN HEAVY CHAIN (B).
FT   DOMAIN     109   187       KRINGLE 1.
FT   DOMAIN     215   292       KRINGLE 2.
FT   DOMAIN     360   617       SERINE PROTEASE.
FT   SITE       200   201       CLEAVAGE (BY THROMBIN).
FT   SITE       323   324       CLEAVAGE (BY FACTOR XA).
FT   SITE       359   360       CLEAVAGE (BY FACTOR XA).
FT   ACT_SITE   402   402       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   458   458       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   564   564       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   MOD_RES     50    50       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     51    51       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     58    58       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     60    60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     63    63       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     64    64       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     69    69       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     70    70       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     73    73       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     76    76       GAMMA-CARBOXYGLUTAMIC ACID.
FT   CARBOHYD   120   120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   144   144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   412   412       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   552   552       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    61    66       BY SIMILARITY.
FT   DISULFID    91   104       BY SIMILARITY.
FT   DISULFID   109   187       BY SIMILARITY.
FT   DISULFID   130   170       BY SIMILARITY.
FT   DISULFID   158   182       BY SIMILARITY.
FT   DISULFID   215   292       BY SIMILARITY.
FT   DISULFID   236   276       BY SIMILARITY.
FT   DISULFID   264   287       BY SIMILARITY.
FT   DISULFID   332   478       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   387   403       BY SIMILARITY.
FT   DISULFID   532   546       BY SIMILARITY.
FT   DISULFID   560   590       BY SIMILARITY.
SQ   SEQUENCE   617 AA;  70411 MW;  AD27D1B71445DB1D CRC64;

  Query Match           100.0%;  Score 69;  DB 1;  Length 617;
  Best Local Similarity 100.0%;  Pred. No. 0.00031;
```

```
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        558 DACEGDSGGPFV 569


RESULT 2
THRB_MOUSE
ID   THRB_MOUSE     STANDARD;      PRT;   618 AA.
AC   P19221;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Prothrombin precursor (EC 3.4.21.5).
GN   F2 OR CF2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Liver;
RX   MEDLINE=91025551; PubMed=2222810;
RA   Friezner Degen S.J., Schaffer L.A., Jamison C.S., Grant S.G.,
RA   Fitzgibbon J.J., Pai J.-A., Chapman V.M., Elliott R.W.;
RT   "Characterization of the cDNA coding for mouse prothrombin and
RT   localization of the gene on mouse chromosome 2.";
RL   DNA Cell Biol. 9:487-498(1990).
RN   [2]
RP   SEQUENCE OF 384-618 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., Macgillivray R.T.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Gly; activates
CC       fibrinogen to fibrin and releases fibrinopeptide A and B.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION
CC       OF PROTHROMBIN TO THROMBIN.
CC   -!- MISCELLANEOUS: PROTHROMBIN IS ACTIVATED ON THE SURFACE OF A
CC       PHOSPHOLIPID MEMBRANE THAT BINDS THE AMINO END OF PROTHROMBIN &
CC       FACTORS VA & XA IN CA-DEPENDENT INTERACTIONS; FACTOR XA REMOVES
CC       THE ACTIVATION PEPTIDE & CLEAVES THE REMAINING PART INTO LIGHT &
CC       HEAVY CHAINS. THE ACTIVATION PROCESS STARTS SLOWLY BECAUSE FACTOR
CC       V ITSELF HAS TO BE ACTIVATED BY THE INITIAL, SMALL AMOUNTS OF
CC       THROMBIN.
CC   -!- MISCELLANEOUS: THROMBIN CAN ITSELF CLEAVE THE AMINO TERMINAL
CC       FRAGMENT (FRAGMENT 1) OF THE PROTHROMBIN, PRIOR TO ITS ACTIVATION
CC       BY FACTOR XA.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   -!- SIMILARITY: Contains 2 kringle domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52308; CAA36548.1; -.
DR   EMBL; M81394; AAA40435.1; -.
DR   PIR; A35827; A35827.
DR   HSSP; P00734; 1B7X.
DR   MEROPS; S01.217; -.
DR   MGD; MGI:88380; F2.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00051; kringle; 2.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00001; GLABLOOD.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   ProDom; PD000395; Kringle; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00130; KR; 2.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00021; KRINGLE_1; 2.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Blood coagulation; Plasma; Calcium-binding; Glycoprotein; Repeat;
KW   Vitamin K; Zymogen; Gamma-carboxyglutamic acid; Acute phase; Liver;
KW   Hydrolase; Serine protease; Kringle; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   PROPEP      25     43
FT   CHAIN        44    618       PROTHROMBIN.
FT   PEPTIDE      44    200       ACTIVATION PEPTIDE (FRAGMENT 1).
FT   PEPTIDE     201    324       ACTIVATION PEPTIDE (FRAGMENT 2).
FT   CHAIN       325    360       THROMBIN LIGHT CHAIN (A).
FT   CHAIN       361    618       THROMBIN HEAVY CHAIN (B).
FT   DOMAIN      109    187       KRINGLE 1.
FT   DOMAIN      215    292       KRINGLE 2.
FT   DOMAIN      361    618       SERINE PROTEASE.
FT   SITE        200    201       CLEAVAGE (BY THROMBIN).
FT   SITE        324    325       CLEAVAGE (BY FACTOR XA).
```

```
FT   SITE        360    361       CLEAVAGE (BY FACTOR XA).
FT   ACT_SITE    403    403       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    459    459       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    565    565       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   MOD_RES      50     50       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      51     51       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      58     58       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      60     60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      63     63       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      64     64       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      69     69       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      70     70       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      73     73       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      76     76       GAMMA-CARBOXYGLUTAMIC ACID.
FT   DISULFID     61     66       BY SIMILARITY.
FT   DISULFID     91    104       BY SIMILARITY.
FT   DISULFID    109    187       BY SIMILARITY.
FT   DISULFID    130    170       BY SIMILARITY.
FT   DISULFID    158    182       BY SIMILARITY.
FT   DISULFID    215    293       BY SIMILARITY.
FT   DISULFID    236    276       BY SIMILARITY.
FT   DISULFID    264    288       BY SIMILARITY.
FT   DISULFID    333    479       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    388    404       BY SIMILARITY.
FT   DISULFID    533    547       BY SIMILARITY.
FT   DISULFID    561    591       BY SIMILARITY.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .).
SQ   SEQUENCE   618 AA;  70268 MW;  B89F719AAFD601E0 CRC64;

  Query Match           100.0%;  Score 69;  DB 1;  Length 618;
  Best Local Similarity  100.0%;  Pred. No. 0.00031;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        559 DACEGDSGGPFV 570


RESULT 3
THRB_HUMAN
ID   THRB_HUMAN     STANDARD;      PRT;   622 AA.
AC   P00734;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Prothrombin precursor (EC 3.4.21.5) (Coagulation factor II).
GN   F2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88077877; PubMed=2825773;
RA   Degen S.J.F., Davie E.W.;
RT   "Nucleotide sequence of the gene for human prothrombin.";
RL   Biochemistry 26:6165-6177(1987).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT MET-165.
RA   Rieder M.J., Armel T.Z., Carrington D.P., Chung M.-W., Lee K.L.,
RA   Ozuna M., Poel C.L., Toth E.J., Yi Q., Nickerson D.A.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 8-622 FROM N.A.
RX   MEDLINE=83231469; PubMed=6305407;
RA   Degen S.J.F., McGillivray R.T.A., Davie E.W.;
RT   "Characterization of the complementary deoxyribonucleic acid and gene
RT   coding for human prothrombin.";
RL   Biochemistry 22:2087-2097(1983).
RN   [4]
RP   SEQUENCE OF 44-314.
RX   MEDLINE=77193964; PubMed=266717;
RA   Walz D.A., Hewett-Emmett D., Seegers W.H.;
RT   "Amino acid sequence of human prothrombin fragments 1 and 2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 74:1969-1972(1977).
RN   [5]
RP   SEQUENCE OF 315-622.
RX   MEDLINE=77207112; PubMed=873923;
RA   Butkowski R.J., Elion J., Downing M.R., Mann K.G.;
RT   "Primary structure of human prethrombin 2 and alpha-thrombin.";
RL   J. Biol. Chem. 252:4942-4957(1977).
RN   [6]
RP   PROCESSING.
RX   MEDLINE=87008532; PubMed=3759958;
RA   Rabiet M.J., Blashill A., Furie B., Furie B.C.;
RT   "Prothrombin fragment 1 X 2 X 3, a major product of prothrombin
RT   activation in human plasma.";
RL   J. Biol. Chem. 261:13210-13215(1986).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=90059942; PubMed=2583108;
RA   Bode W., Mayr I., Baumann U., Huber R., Stone S.R., Hofsteenge J.;
RT   "The refined 1.9 A crystal structure of human alpha-thrombin:
RT   interaction with D-Phe-Pro-Arg chloromethylketone and significance of
RT   the Tyr-Pro-Pro-Trp insertion segment.";
RL   EMBO J. 8:3467-3475(1989).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=90327074; PubMed=2374926;
RA   Rydel T.J., Ravichandran K.G., Tulinsky A., Bode W., Huber R.,
RA   Roitsch C., Fenton J.W. II;
RT   "The structure of a complex of recombinant hirudin and human alpha-
RT   thrombin.";
RL   Science 249:277-280(1990).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=94350942; PubMed=8071320;
RA   Rydel T.J., Yin M., Padmanabhan K.P., Blankenship D.T., Cardin A.D.,
RA   Correa P.E., Fenton J.W. II, Tulinsky A.;
RT   "Crystallographic structure of human gamma-thrombin.";
RL   J. Biol. Chem. 269:22000-22006(1994).
```

RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=97357286; PubMed=9214615;
RA van de Locht A., Bode W., Huber R., le Bonniec B.F., Stone S.R.,
RA Esmon C.T., Stubbs M.T.;
RT "The thrombin E192Q-BPTI complex reveals gross structural
RT rearrangements: implications for the interaction with antithrombin
RT and thrombomodulin.";
RL EMBO J. 16:2977-2984(1997).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 328-601.
RX MEDLINE=99162521; PubMed=10051558;
RA Guinto E.R., Caccia S., Rose T., Fuetterer K., Waksman G., di Cera E.;
RT "Unexpected crucial role of residue 225 in serine proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1852-1857(1999).
RN [12]
RP VARIANT BARCELONA.
RX MEDLINE=87033739; PubMed=3771562;
RA Rabiet M.-J., Furie B.C., Furie B.;
RT "Molecular defect of prothrombin Barcelona. Substitution of cysteine
RT for arginine at residue 273.";
RL J. Biol. Chem. 261:15045-15048(1986).
RN [13]
RP VARIANT FRANKFURT.
RX MEDLINE=95313001; PubMed=7792730;
RA Degen S.J.F., McDowell S.A., Sparks L.M., Scharrer I.;
RT "Prothrombin Frankfurt: a dysfunctional prothrombin characterized by
RT substitution of Glu-466 by Ala.";
RL Thromb. Haemost. 73:203-209(1995).
RN [14]
RP VARIANTS HIMI-1 AND HIMI-2.
RX MEDLINE=93043342; PubMed=1421398;
RA Morishita E., Saito M., Kumabashiri I., Asakura H., Matsuda T.,
RA Yamaguchi K.;
RT "Prothrombin Himi: a compound heterozygote for two dysfunctional
RT prothrombin molecules (Met-337-->Thr and Arg-388-->His).";
RL Blood 80:2275-2280(1992).
RN [15]
RP VARIANT PADUA-1.
RX MEDLINE=95169898; PubMed=7865694;
RA James H.L., Kim D.J., Zheng D.-Q., Girolami A.;
RT "Prothrombin Padua I: incomplete activation due to an amino acid
RT substitution at a factor Xa cleavage site.";
RL Blood Coagul. Fibrinolysis 5:841-844(1994).
RN [16]
RP VARIANT QUICK-1.
RX MEDLINE=89207504; PubMed=3242619;
RA Henriksen R.A., Mann K.G.;
RT "Identification of the primary structural defect in the dysthrombin
RT thrombin Quick I: substitution of cysteine for arginine-382.";
RL Biochemistry 27:9160-9165(1988).
RN [17]
RP VARIANT QUICK-2.
RX MEDLINE=89247398; PubMed=2719946;
RA Henriksen R.A., Mann K.G.;
RT "Substitution of valine for glycine-558 in the congenital dysthrombin
RT thrombin Quick II alters primary substrate specificity.";
RL Biochemistry 28:2078-2082(1989).
RN [18]
RP VARIANT SALAKTA.
RX MEDLINE=92378975; PubMed=1354985;
RA Miyata T., Aruga R., Umeyama H., Bezeaud A., Guillin M.-C.,
RA Iwanaga S.;
RT "Prothrombin Salakta: substitution of glutamic acid-466 by alanine
RT reduces the fibrinogen clotting activity and the esterase activity.";
RL Biochemistry 31:7457-7462(1992).
RN [19]
RP VARIANT TOKUSHIMA.
RX MEDLINE=87185407; PubMed=3567158;
RA Miyata T., Morita T., Inomoto T., Kawauchi S., Shirakami A.,
RA Iwanaga S.;
RT "Prothrombin Tokushima, a replacement of arginine-418 by tryptophan
RT that impairs the fibrinogen clotting activity of derived thrombin
RT Tokushima.";
RL Biochemistry 26:1117-1122(1987).
RN [20]
RP VARIANT TOKUSHIMA.
RX MEDLINE=87101511; PubMed=3801671;
RA Inomoto T., Shirakami A., Kawauchi S., Shigekiyo T., Saito S.,
RA Miyoshi K., Morita T., Iwanaga S.;
RT "Prothrombin Tokushima: characterization of dysfunctional thrombin
RT derived from a variant of human prothrombin.";
RL Blood 69:565-569(1987).
RN [21]
RP VARIANT TOKUSHIMA.
RX MEDLINE=92256895; PubMed=1349838;
RA Iwahana H., Yoshimoto K., Shigekiyo T., Shirakami A., Saito S.,
RA Itakura M.;
RT "Detection of a single base substitution of the gene for prothrombin
RT Tokushima. The application of PCR-SSCP for the genetic and molecular
RT analysis of dysprothrombinemia.";
RL Int. J. Hematol. 55:93-100(1992).
RN [22]
RP VARIANT TYPE-3.
RX MEDLINE=83204687; PubMed=6405779;
RA Board P.G., Shaw D.C.;
RT "Determination of the amino acid substitution in human prothrombin
RT type 3 (157 Glu leads to Lys) and the localization of a third
RT thrombin cleavage site.";
RL Br. J. Haematol. 54:245-254(1983).
RN [23]
RP VARIANTS MET-165 AND THR-386.
RX MEDLINE=99318093; PubMed=10391209;
RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA Lander E.S.;
RT "Characterization of single-nucleotide polymorphisms in coding regions
RT of human genes.";
RL Nat. Genet. 22:231-238(1999).
RN [24]
RP ERRATUM.
RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,

```
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RL   Nat. Genet. 23:373-373(1999).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Gly; activates
CC       fibrinogen to fibrin and releases fibrinopeptide A and B.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION

  Query Match           100.0%;  Score 69;  DB 1;  Length 622;
  Best Local Similarity  100.0%;  Pred. No. 0.00031;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        562 DACEGDSGGPFV 573


RESULT 4
THRB_BOVIN
ID   THRB_BOVIN     STANDARD;      PRT;   625 AA.
AC   P00735;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Prothrombin precursor (EC 3.4.21.5).
GN   F2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88245190; PubMed=3379642;
RA   Irwin D.M., Robertson K.A., Macgillivray R.T.A.;
RT   "Structure and evolution of the bovine prothrombin gene.";
RL   J. Mol. Biol. 200:31-45(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84203525; PubMed=6326805;
RA   McGillivray R.T.A., Davie E.W.;
RT   "Characterization of bovine prothrombin mRNA and its translation
RT   product.";
RL   Biochemistry 23:1626-1634(1984).
RN   [3]
RP   SEQUENCE OF 44-625, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RA   Magnusson S., Sottrup-Jensen L., Petersen T.E., Claeys H.;
RL   (In) Hemker H.C., Veltkamp J.J. (eds.);
RL   Boerhaave symposium on prothrombin and related coagulation factors,
RL   pp.25-46, Leiden University Press, Leiden (1975).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE=86296631; PubMed=3741841;
RA   Park C.H., Tulinsky A.;
RT   "Three-dimensional structure of the kringle sequence: structure of
RT   prothrombin fragment 1.";
RL   Biochemistry 25:3977-3982(1986).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE=91311686; PubMed=1856869;
RA   Seshadri T.-P., Tulinsky A., Skrzypczak-Jankun E., Park C.H.;
RT   "Structure of bovine prothrombin fragment 1 refined at 2.25-A
RT   resolution.";
RL   J. Mol. Biol. 220:481-494(1991).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE=92190185; PubMed=1547238;
RA   Soriano-Garcia M., Padmanabhan K., de Vos A.M., Tulinsky A.;
RT   "The Ca2+ ion and membrane binding structure of the Gla domain of Ca-
RT   prothrombin fragment 1.";
RL   Biochemistry 31:2554-2566(1992).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=92218459; PubMed=1560020;
RA   Martin P.D., Robertson W., Turk D., Huber R., Bode W., Edwards B.F.P.;
RT   "The structure of residues 7-16 of the A alpha-chain of human
RT   fibrinogen bound to bovine thrombin at 2.3-A resolution.";
RL   J. Biol. Chem. 267:7911-7920(1992).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=92389319; PubMed=1518046;
RA   Brandstetter H., Turk D., Hoeffken H.W., Grosse D., Stuerzebecher J.,
RA   Martin P.D., Edwards B.F.P., Bode W.;
RT   "Refined 2.3 A X-ray crystal structure of bovine thrombin complexes
RT   formed with the benzamidine and arginine-based thrombin inhibitors
RT   NAPAP, 4-TAPAP and MQPA. A starting point for improving
RT   antithrombotics.";
RL   J. Mol. Biol. 226:1085-1089(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF COMPLEX WITH ORNITHODORIN.
RX   MEDLINE=97102783; PubMed=8947023;
RA   van de Locht A., Stubbs M.T., Bode W., Friedrich T., Bollschweiler C.,
RA   Hoffken W., Huber R.;
RT   "The ornithodorin-thrombin crystal structure, a key to the TAP
RT   enigma?";
RL   EMBO J. 15:6011-6017(1996).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH TRIABIN.
RX   MEDLINE=98004486; PubMed=9342325;
RA   Fuentes-Prior P., Noeske-Jungblut C., Donner P., Schleuning W.D.,
RA   Huber R., Bode W.;
RT   "Structure of the thrombin complex with triabin, a lipocalin-like
RT   exosite-binding inhibitor derived from a triatomine bug.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11845-11850(1997).
RN   [11]
RP   GENE STRUCTURE.
```

```
RX   MEDLINE=86077733; PubMed=3000440;
RA   Irwin D.M., Ahern K.G., Pearson G.D., McGillivray R.T.A.;
RT   "Characterization of the bovine prothrombin gene.";
RL   Biochemistry 24:6854-6861(1985).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Gly; activates
CC       fibrinogen to fibrin and releases fibrinopeptide A and B.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION
CC       OF PROTHROMBIN TO THROMBIN.
CC   -!- MISCELLANEOUS: PROTHROMBIN IS ACTIVATED ON THE SURFACE OF A
CC       PHOSPHOLIPID MEMBRANE THAT BINDS THE AMINO END OF PROTHROMBIN &
CC       FACTORS VA & XA IN CA-DEPENDENT INTERACTIONS; FACTOR XA REMOVES
CC       THE ACTIVATION PEPTIDE & CLEAVES THE REMAINING PART INTO LIGHT &
CC       HEAVY CHAINS. THE ACTIVATION PROCESS STARTS SLOWLY BECAUSE FACTOR
CC       V ITSELF HAS TO BE ACTIVATED BY THE INITIAL, SMALL AMOUNTS OF
CC       THROMBIN.
CC   -!- MISCELLANEOUS: THROMBIN CAN ITSELF CLEAVE THE AMINO TERMINAL
CC       FRAGMENT (FRAGMENT 1) OF THE PROTHROMBIN, PRIOR TO ITS ACTIVATION
CC       BY FACTOR XA.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   -!- SIMILARITY: Contains 2 kringle domains.
CC   -!- DATABASE: NAME=ProZyme technical fact sheet;
CC       WWW="http://www.prozyme.com/technical/thrombindata.html".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; V00135; CAA23451.1; -.
DR   EMBL; J00041; AAA30781.1; -.
DR   PIR; S02537; TBBO.
DR   PDB; 1BBR; 31-JAN-94.
DR   PDB; 1ETR; 31-JAN-94.
DR   PDB; 1ETS; 31-JAN-94.
DR   PDB; 1ETT; 31-JAN-94.
DR   PDB; 1HRT; 31-JAN-94.
DR   PDB; 2PF1; 31-JAN-94.
DR   PDB; 2PF2; 31-JAN-94.
DR   PDB; 2SPT; 31-MAY-94.
DR   PDB; 1MKW; 07-JUL-97.
DR   PDB; 1MKX; 07-JUL-97.
DR   PDB; 1TBQ; 14-OCT-96.
DR   PDB; 1TBR; 14-OCT-96.
DR   PDB; 1TOC; 23-JUL-97.
DR   PDB; 1VIT; 21-APR-97.
DR   PDB; 1YCP; 06-MAY-98.
DR   PDB; 1A0H; 17-JUN-98.
DR   PDB; 1AVG; 16-FEB-99.
DR   PDB; 1BTH; 24-DEC-97.
DR   PDB; 1ID5; 12-SEP-01.
DR   PDB; 1UVT; 19-NOV-97.
DR   PDB; 2HPP; 31-JAN-94.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00051; kringle; 2.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00001; GLABLOOD.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   ProDom; PD000395; Kringle; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00130; KR; 2.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00021; KRINGLE_1; 2.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Blood coagulation; Plasma; Calcium-binding; Glycoprotein; Repeat;
KW   Vitamin K; Zymogen; Gamma-carboxyglutamic acid; Acute phase; Liver;
KW   Hydrolase; Serine protease; Kringle; Signal; 3D-structure.
FT   SIGNAL       1    24       POTENTIAL.
FT   PROPEP       25   43
FT   CHAIN        44   625      PROTHROMBIN.
FT   PEPTIDE      44   199      ACTIVATION PEPTIDE (FRAGMENT 1).
FT   PEPTIDE      200  317      ACTIVATION PEPTIDE (FRAGMENT 2).
FT   CHAIN        318  366      THROMBIN LIGHT CHAIN (A).
FT   CHAIN        367  625      THROMBIN HEAVY CHAIN (B).
FT   DOMAIN       109  187      KRINGLE 1.
FT   DOMAIN       214  292      KRINGLE 2.
FT   DOMAIN       367  625      SERINE PROTEASE.
FT   SITE         199  200      CLEAVAGE (BY THROMBIN).
FT   SITE         317  318      CLEAVAGE (BY FACTOR XA).
FT   SITE         366  367      CLEAVAGE (BY FACTOR XA).
FT   ACT_SITE     409  409      CHARGE RELAY SYSTEM.
FT   ACT_SITE     465  465      CHARGE RELAY SYSTEM.
FT   ACT_SITE     571  571      CHARGE RELAY SYSTEM.
FT   MOD_RES      50   50       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      51   51       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      58   58       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      60   60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      63   63       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      64   64       GAMMA-CARBOXYGLUTAMIC ACID.
```

```
  Query Match             100.0%;  Score 69;  DB 1;  Length 625;
  Best Local Similarity   100.0%;  Pred. No. 0.00032;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        565 DACEGDSGGPFV 576


RESULT 5
HEPS_HUMAN
ID   HEPS_HUMAN     STANDARD;      PRT;   417 AA.
AC   P05981;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Serine protease hepsin (EC 3.4.21.-) (Transmembrane protease, serine
DE   1).
GN   HPN OR TMPRSS1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=88209431; PubMed=2835076;
RA   Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
RT   "A novel trypsin-like serine protease (hepsin) with a putative
RT   transmembrane domain expressed by human liver and hepatoma cells.";
RL   Biochemistry 27:1067-1074(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas, and Spleen;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=91358502; PubMed=1885621;
RA   Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
RA   Chou S.H., Kurachi K.;
RT   "Hepsin, a cell membrane-associated protease. Characterization,
RT   tissue distribution, and gene localization.";
RL   J. Biol. Chem. 266:16948-16953(1991).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=93348237; PubMed=8346233;
RA   Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
RT   "Hepsin, a putative cell-surface serine protease, is required for
RT   mammalian cell growth.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
CC   -!- FUNCTION: Plays an essential role in cell growth and maintenance
CC       of cell morphology.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- TISSUE SPECIFICITY: Present in most tissues, with the highest
CC       level in liver.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M18930; AAA36013.1; -.
DR   EMBL; X07732; CAA30558.1; -.
DR   EMBL; X07002; CAA30058.1; -.
DR   EMBL; BC025716; AAH25716.1; -.
DR   PIR; S00845; S00845.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.224; -.
DR   Genew; HGNC:5155; HPN.
DR   MIM; 142440; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0008151; P:cell growth and/or maintenance; TAS.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT   CHAIN         1    162       SERINE PROTEASE HEPSIN, NON-CATALYTIC
FT                                CHAIN (POTENTIAL).
FT   CHAIN       163    417       SERINE PROTEASE HEPSIN, CATALYTIC CHAIN
FT                                (POTENTIAL).
FT   DOMAIN        1     17       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     18     44       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       45    417       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      163    417       SERINE PROTEASE.
```

```
FT   ACT_SITE    203    203       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    257    257       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    353    353       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    153    277       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    188    204       BY SIMILARITY.
FT   DISULFID    322    338       BY SIMILARITY.
FT   DISULFID    349    381       BY SIMILARITY.
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   417 AA;  45011 MW;  B2086FF661E551D7 CRC64;

  Query Match             95.7%;  Score 66;  DB 1;  Length 417;
  Best Local Similarity   91.7%;  Pred. No. 0.00067;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              |||:|||||||||
Db        347 DACQGDSGGPFV 358


RESULT 6
HEPS_MOUSE
ID   HEPS_MOUSE     STANDARD;      PRT;   436 AA.
AC   035453; Q9CW97;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Serine protease hepsin (EC 3.4.21.-).
GN   HPN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Liver;
RX   MEDLINE=98058912; PubMed=9395459;
RA   Vu T.-K.H., Liu R.W., Haaksma C., Tomasek J.J., Howard E.W.;
RT   "Identification and cloning of the membrane-associated serine
RT   protease, hepsin, from mouse preimplantation embryos.";
RL   J. Biol. Chem. 272:31315-31320(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=99339944; PubMed=10411637;
RA   Kawamura S., Kurachi S., Deyashiki Y., Kurachi K.;
RT   "Complete nucleotide sequence, origin of isoform and functional
RT   characterization of the mouse hepsin gene.";
RL   Eur. J. Biochem. 262:755-764(1999).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: Plays an essential role in cell growth and maintenance
CC       of cell morphology.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1; Synonyms=1a;
CC         IsoId=035453-1; Sequence=Displayed;
CC         Note=Minor isoform;
CC       Name=2; Synonyms=2a;
CC         IsoId=035453-2; Sequence=VSP_007232;
CC         Note=Major isoform;
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to
CC       frameshifts in positions 155, 191 and 233.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF030065; AAB84221.1; -.
DR   EMBL; AK002694; BAB22289.2; ALT_FRAME.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.224; -.
DR   MGD; MGI:1196620; Hpn.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR001190; Srcr_receptor.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00202; SR; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW   Alternative splicing.
FT   CHAIN         1    181       SERINE PROTEASE HEPSIN, NON-CATALYTIC
FT                                CHAIN (POTENTIAL).
FT   CHAIN       182    436       SERINE PROTEASE HEPSIN, CATALYTIC CHAIN
```

```
FT                                     (POTENTIAL).
FT   DOMAIN       21     36       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     37     63       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                     (POTENTIAL).
FT   DOMAIN       64    436       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      182    436       SERINE PROTEASE.
FT   ACT_SITE    222    222       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    276    276       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    372    372       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    172    296       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    207    223       BY SIMILARITY.
FT   DISULFID    341    357       BY SIMILARITY.
FT   DISULFID    368    400       BY SIMILARITY.
FT   CARBOHYD    131    131       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     25     44       Missing (in isoform 2).
FT                                /FTId=VSP_007232.
FT   CONFLICT     85     85       L -> F (IN REF. 2 AND 3).
FT   CONFLICT    204    204       T -> Y (IN REF. 3).
FT   CONFLICT    214    214       G -> R (IN REF. 3).
FT   CONFLICT    228    229       NR -> ET (IN REF. 3).
FT   CONFLICT    264    264       P -> L (IN REF. 3).
FT   CONFLICT    281    281       H -> N (IN REF. 3).
SQ   SEQUENCE   436 AA;  46787 MW;  4A0993148C620BD0 CRC64;

  Query Match             95.7%;  Score 66;  DB 1;  Length 436;
  Best Local Similarity   91.7%;  Pred. No. 0.0007;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              |||:||||||||
Db        366 DACQGDSGGPFV 377

Search completed: February 11, 2004, 14:54:03
Job time : 5.03226 secs
```

```
                                GenCore version 5.1.6
                          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         February 11, 2004, 14:47:57 ; Search time 20.5161 Seconds
                                             (without alignments)
                                             150.936 Million cell updates/sec

Title:          US-10-050-611-2
Perfect score:  69
Sequence:       1 DACEGDSGGPFV 12

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        SPTREMBL_23:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                    SUMMARIES

                          %
Result                    Query
  No.    Score            Match  Length  DB  ID                      Description
 ---------------------------------------------------------------------------
```

1	69	100.0	235	6	Q28731	Q28731 oryctolagus
2	69	100.0	235	13	Q91004	Q91004 gecko gecko
3	69	100.0	235	13	Q90387	Q90387 cynops pyrr
4	69	100.0	239	13	Q91218	Q91218 oncorhynchu
5	69	100.0	607	13	Q91001	Q91001 gallus gall
6	69	100.0	608	13	Q9PTW7	Q9ptw7 struthio ca
7	66	95.7	234	13	Q90244	Q90244 acipenser t
8	66	95.7	435	11	Q9CW97	Q9cw97 mus musculu
9	66	95.7	1524	13	Q91674	Q91674 xenopus lae
10	64	92.8	420	13	Q90504	Q90504 eptatretus
11	63	91.3	195	4	Q8J008	Q8j008 homo sapien
12	63	91.3	195	4	Q8J007	Q8j007 homo sapien
13	63	91.3	195	4	Q8J006	Q8j006 homo sapien
14	63	91.3	195	4	Q8IXB4	Q8ixb4 homo sapien
15	63	91.3	211	4	Q8J009	Q8j009 homo sapien
16	63	91.3	255	5	Q9NBC9	Q9nbc9 glossina mo
17	63	91.3	257	11	Q8BZ04	Q8bz04 mus musculu
18	63	91.3	267	5	Q9BK47	Q9bk47 luidia foli
19	63	91.3	358	5	O45029	O45029 drosophila
20	63	91.3	371	5	Q8MRY3	Q8mry3 drosophila
21	63	91.3	417	11	Q8BZ10	Q8bz10 mus musculu
22	63	91.3	456	6	Q9TTR0	Q9ttr0 canis famil
23	63	91.3	974	13	Q90WD8	Q90wd8 bufo japoni
24	63	91.3	1374	5	Q9VSU0	Q9vsu0 drosophila
25	63	91.3	1449	5	Q9U1I2	Q9u1i2 drosophila
26	63	91.3	1450	5	Q8IQB8	Q8iqb8 drosophila
27	63	91.3	1462	5	Q9U1I3	Q9u1i3 drosophila
28	63	91.3	2382	5	Q9BI19	Q9bi19 drosophila
29	63	91.3	2409	5	Q960G6	Q960g6 drosophila
30	63	91.3	2786	5	Q9VSU2	Q9vsu2 drosophila
31	61	88.4	248	5	Q8IRE2	Q8ire2 drosophila
32	60	87.0	85	5	Q8MVL1	Q8mvl1 boltenia vi
33	60	87.0	155	5	Q9Y1K4	Q9y1k4 anopheles g
34	60	87.0	187	5	O45045	O45045 scirpophaga
35	60	87.0	200	11	Q924U6	Q924u6 mus musculu
36	60	87.0	234	11	Q9CV76	Q9cv76 mus musculu
37	60	87.0	247	13	Q9W7Q5	Q9w7q5 paralichthy
38	60	87.0	250	5	Q9V514	Q9v514 drosophila
39	60	87.0	252	5	O76498	O76498 diaprepes a
40	60	87.0	253	5	Q8SXZ4	Q8sxz4 drosophila
41	60	87.0	253	5	Q8MKZ1	Q8mkz1 drosophila
42	60	87.0	254	5	Q9XYY0	Q9xyy0 rhyzopertha
43	60	87.0	254	5	O76520	O76520 stomoxys ca
44	60	87.0	254	11	Q8CGR4	Q8cgr4 mus musculu
45	60	87.0	255	3	Q9Y7A9	Q9y7a9 metarhizium

ALIGNMENTS

```
RESULT 1
Q28731
ID   Q28731      PRELIMINARY;      PRT;   235 AA.
AC   Q28731;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Thrombin (Fragment).
GN   THROMBIN.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., MacGillivray R.T.A.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   Amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
DR   EMBL; M81396; AAA31477.1; -.
DR   HSSP; P00734; 1UVS.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Protease; Serine protease.
FT   NON_TER       1       1
SQ   SEQUENCE   235 AA;  27093 MW;  92FF3E4F93B360E0 CRC64;

  Query Match           100.0%;  Score 69;  DB 6;  Length 235;
  Best Local Similarity  100.0%;  Pred. No. 0.00086;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        176 DACEGDSGGPFV 187


RESULT 2
Q91004
ID   Q91004      PRELIMINARY;      PRT;   235 AA.
AC   Q91004;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Thrombin (Fragment).
GN   THROMBIN.
OS   Gecko gecko (Tokay gecko).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Gekkota; Gekkonidae; Gekko.
OX   NCBI_TaxID=36310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
```

```
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., MacGillivray R.T.A.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   Amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
DR   EMBL; M81392; AAA49309.1; -.
DR   HSSP; P00734; 1B7X.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Protease; Serine protease.
FT   NON_TER       1      1
SQ   SEQUENCE   235 AA;  26933 MW;  122A5C09F6F2276A CRC64;

  Query Match             100.0%;  Score 69;  DB 13;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 0.00086;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        175 DACEGDSGGPFV 186


RESULT 3
Q90387
ID   Q90387      PRELIMINARY;      PRT;   235 AA.
AC   Q90387;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Thrombin (Fragment).
GN   THROMBIN.
OS   Cynops pyrrhogaster (Japanese common newt).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX   NCBI_TaxID=8330;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., MacGillivray R.T.A.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   Amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
DR   EMBL; M81395; AAA49391.1; -.
DR   HSSP; P00734; 1UVS.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Protease; Serine protease.
FT   NON_TER       1      1
SQ   SEQUENCE   235 AA;  27272 MW;  49264DD29A57A41F CRC64;

  Query Match             100.0%;  Score 69;  DB 13;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 0.00086;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        175 DACEGDSGGPFV 186


RESULT 4
Q91218
ID   Q91218      PRELIMINARY;      PRT;   239 AA.
AC   Q91218;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Thrombin (Fragment).
GN   THROMBIN.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., MacGillivray R.T.A.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   Amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
DR   EMBL; M81398; AAA49433.1; -.
DR   HSSP; P00734; 1B7X.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
```

```
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Protease; Serine protease.
FT   NON_TER       1      1
SQ   SEQUENCE   239 AA;  27396 MW;  F0F43F9A3205BF38 CRC64;

  Query Match             100.0%;  Score 69;  DB 13;  Length 239;
  Best Local Similarity   100.0%;  Pred. No. 0.00087;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        175 DACEGDSGGPFV 186


RESULT 5
Q91001
ID   Q91001      PRELIMINARY;      PRT;   607 AA.
AC   Q91001;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Thrombin.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., MacGillivray R.T.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94223694; PubMed=7513365;
RA   Banfield D.K., Irwin D.M., Walz D.A., MacGillivray R.T.;
RT   "Evolution of prothrombin: isolation and characterization of the cDNAs
RT   encoding chicken and hagfish prothrombin.";
RL   J. Mol. Evol. 38:177-187(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Banfield D.K.;
RL   Submitted (DEC-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR   EMBL; M81391; AAA21619.1; -.
DR   HSSP; P00734; 1UVS.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00051; kringle; 2.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00001; GLABLOOD.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   ProDom; PD000395; Kringle; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00130; KR; 2.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00021; KRINGLE_1; 2.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Glycoprotein; Hydrolase; Kringle; Protease; Serine protease.
SQ   SEQUENCE   607 AA;  69110 MW;  002F3606EA36270F CRC64;

  Query Match             100.0%;  Score 69;  DB 13;  Length 607;
  Best Local Similarity   100.0%;  Pred. No. 0.0022;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        548 DACEGDSGGPFV 559


RESULT 6
Q9PTW7
ID   Q9PTW7      PRELIMINARY;      PRT;   608 AA.
AC   Q9PTW7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Prothrombin.
GN   OSPT.
OS   Struthio camelus (Ostrich).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;
OC   Struthio.
OX   NCBI_TaxID=8801;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20579470; PubMed=11137455;
RA   Frost C., Naude R., Oelofsen W., Muramoto K., Naganuma T., Ogawa T.;
RT   "Purification and characterization of ostrich prothrombin.";
RL   Int. J. Biochem. Cell Biol. 32:1151-1159(2000).
CC   -!- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR   EMBL; AB028871; BAA89046.1; -.
DR   HSSP; P00734; 1UVS.
DR   MEROPS; S01.217; -.
```

```
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00051; kringle; 2.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00001; GLABLOOD.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   ProDom; PD000395; Kringle; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00130; KR; 2.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00021; KRINGLE_1; 2.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Glycoprotein; Hydrolase; Kringle; Protease; Serine protease.
SQ   SEQUENCE   608 AA;  69392 MW;  11B974B9AEE54EA2 CRC64;

  Query Match             100.0%;  Score 69;  DB 13;  Length 608;
  Best Local Similarity   100.0%;  Pred. No. 0.0022;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              ||||||||||||
Db        548 DACEGDSGGPFV 559


RESULT 7
Q90244
ID   Q90244      PRELIMINARY;      PRT;   234 AA.
AC   Q90244;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Thrombin (Fragment).
GN   THROMBIN.
OS   Acipenser transmontanus (White sturgeon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;
OC   Acipenser.
OX   NCBI_TaxID=7904;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., MacGillivray R.T.A.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   Amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
DR   EMBL; M81399; AAA48514.1; -.
DR   HSSP; P00734; 2HNT.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Protease; Serine protease.
FT   NON_TER       1      1
SQ   SEQUENCE   234 AA;  26846 MW;  45C558D6618E0585 CRC64;

  Query Match             95.7%;  Score 66;  DB 13;  Length 234;
  Best Local Similarity   91.7%;  Pred. No. 0.0027;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              |:|||||||||||
Db        175 DSCEGDSGGPFV 186


RESULT 8
Q9CW97
ID   Q9CW97      PRELIMINARY;      PRT;   435 AA.
AC   Q9CW97;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Adult male kidney cDNA, RIKEN full-length enriched library,
DE   clone:0610030A17 product:hepsin, full insert sequence.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=21085660; PubMed=11217851;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=99279253; PubMed=10349636;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=20499374; PubMed=11042159;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=20530913; PubMed=11076861;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
DR   EMBL; AK002694; BAB22289.2; -.
SQ   SEQUENCE   435 AA;  45944 MW;  019B2A9DE3EBEF40 CRC64;

  Query Match             95.7%;  Score 66;  DB 11;  Length 435;
  Best Local Similarity   91.7%;  Pred. No. 0.0051;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              |||:|||||||||
Db        365 DACQGDSGGPFV 376

RESULT 9
```

```
Q91674
ID   Q91674      PRELIMINARY;      PRT;  1524 AA.
AC   Q91674;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polyprotein.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99432219; PubMed=10500163;
RA   Lindsay L.L., Yang J.C., Hedrick J.L.;
RT   "Ovochymase, a Xenopus laevis egg extracellular protease, is
RT   translated as part of an unusual polyprotease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:11253-11258(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yang J.C., Lindsay L.L., Hedrick J.L.;
RT   "cDNA Cloning of Ovochymase, a Chymotrypsin-like Protease Released
RT   From Xenopus laevis Eggs at Fertilization.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   -!- SIMILARITY: CONTAINS 4 CUB DOMAINS.
DR   EMBL; U81290; AAC24717.1; -.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.022; -.
DR   MEROPS; S01.245; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000859; CUB_domain.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00431; CUB; 5.
DR   Pfam; PF00089; trypsin; 3.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00042; CUB; 4.
DR   SMART; SM00020; Tryp_SPc; 3.
DR   PROSITE; PS01180; CUB; 5.
DR   PROSITE; PS50240; TRYPSIN_DOM; 3.
DR   PROSITE; PS00134; TRYPSIN_HIS; 3.
DR   PROSITE; PS00135; TRYPSIN_SER; 3.
KW   Hydrolase; Protease; Serine protease.
FT   CHAIN        57    308       SERINE PROTEASE.
FT   CHAIN       584    817       SERINE PROTEASE.
FT   CHAIN      1295   1524       OVOCHYMASE.
SQ   SEQUENCE   1524 AA;  167566 MW;  32EFE42128F37268 CRC64;

  Query Match             95.7%;  Score 66;  DB 13;  Length 1524;
  Best Local Similarity   91.7%;  Pred. No. 0.018;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              |||:|||||||||
Db        241 DACQGDSGGPFV 252
```

```
RESULT 10
Q90504
ID   Q90504      PRELIMINARY;      PRT;   420 AA.
AC   Q90504;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Thrombin.
OS   Eptatretus stoutii (Pacific hagfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniformes;
OC   Myxinidae; Eptatretinae; Eptatretus.
OX   NCBI_TaxID=7765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92212913; PubMed=1557383;
RA   Banfield D.K., MacGillivray R.T.;
RT   "Partial characterization of vertebrate prothrombin cDNAs:
RT   amplification and sequence analysis of the B chain of thrombin from
RT   nine different species.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94223694; PubMed=7513365;
RA   Banfield D.K., Irwin D.M., Walz D.A., MacGillivray R.T.;
RT   "Evolution of prothrombin: isolation and characterization of the cDNAs
RT   encoding chicken and hagfish prothrombin.";
RL   J. Mol. Evol. 38:177-187(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Banfield D.K.;
RL   Submitted (DEC-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR   EMBL; M81393; AAA21620.1; -.
DR   HSSP; P00734; 1UVS.
DR   MEROPS; S01.217; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00051; kringle; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   ProDom; PD000395; Kringle; 1.
DR   SMART; SM00130; KR; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00021; KRINGLE_1; 1.
DR   PROSITE; PS50070; KRINGLE_2; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Kringle; Protease; Serine protease.

SQ   SEQUENCE   420 AA;  47888 MW;  64522AA21A57B67A CRC64;

  Query Match             92.8%;  Score 64;  DB 13;  Length 420;
  Best Local Similarity   91.7%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DACEGDSGGPFV 12
              | ||||||||||
Db        359 DPCEGDSGGPFV 370

Search completed: February 11, 2004, 14:56:04
Job time : 22.5161 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:35:52 ; Search time 49.7097 Seconds
(without alignments)
73.441 Million cell updates/sec

Title: US-10-050-611-3
Perfect score: 131
Sequence: 1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	131	100.0	23	20	AAW83414	Cell growth/adhesi
2	131	100.0	23	21	AAB12893	Nerve tissue regen
3	131	100.0	23	22	AAB70363	Human thrombin rec
4	131	100.0	23	23	AAE22563	Human thrombin hig
5	131	100.0	23	23	AAE20159	Human thrombin pep
6	131	100.0	23	23	AAM50858	Thrombin-derived p
7	131	100.0	116	20	AAW99115	Human zeta 2 preth
8	131	100.0	259	18	AAW11545	Human thrombin Asn
9	131	100.0	259	24	ABP60563	Human thrombin var
10	131	100.0	259	24	ABP60565	Human thrombin var
11	131	100.0	295	16	AAR74775	Wild-type thrombin
12	131	100.0	295	16	AAR74776	Mutant thrombin K5
13	131	100.0	295	16	AAR74777	Mutant thrombin E2
14	131	100.0	295	16	AAR74778	Mutant thrombin E2
15	131	100.0	295	16	AAR74779	Mutant thrombin E2
16	131	100.0	295	16	AAR74780	Mutant thrombin E2
17	131	100.0	295	16	AAR76033	Mutant thrombin E2
18	131	100.0	295	16	AAR76034	Mutant thrombin R2
19	131	100.0	295	16	AAR76035	Mutant thrombin R2
20	131	100.0	295	16	AAR76036	Mutant thrombin R2
21	131	100.0	295	16	AAR76037	Mutant thrombin W5
22	131	100.0	295	16	AAR76038	Mutant thrombin K5
23	131	100.0	295	16	AAR76039	Mutant thrombin W5
24	131	100.0	295	16	AAR76040	Mutant thrombin W5
25	131	100.0	295	18	AAW22892	Human mature throm
26	131	100.0	295	21	AAB08633	Amino acid sequenc
27	131	100.0	295	24	ABP60562	Human thrombin var
28	131	100.0	295	24	ABP60564	Human thrombin var
29	131	100.0	308	20	AAW99109	Human prethrombin
30	131	100.0	376	14	AAR41797	CD4/Thrombin fusio
31	131	100.0	376	20	AAY42789	Human CD4/thrombin
32	131	100.0	376	23	AAU10703	Human CD4-thrombin
33	131	100.0	579	14	AAR35763	Prothrombin (PT).
34	131	100.0	579	18	AAW11546	Human prothrombin
35	131	100.0	579	18	AAW11544	Human prothrombin
36	131	100.0	579	20	AAW99108	Human prothrombin.
37	131	100.0	615	14	AAR38741	Human prothrombin.
38	131	100.0	615	17	AAR96216	Human prothrombin.
39	131	100.0	615	17	AAR90377	Human prothrombin.
40	131	100.0	622	18	AAW11543	Human preprothromb
41	131	100.0	622	20	AAY49566	Platelet membrane
42	131	100.0	622	24	ABG74671	Human F2 protein.
43	124	94.7	111	20	AAW99113	Bovine zeta 2 pret
44	124	94.7	308	20	AAW99107	Bovine prethrombin
45	124	94.7	582	20	AAW99106	Bovine prothrombin

ALIGNMENTS

RESULT 1
AAW83414
ID AAW83414 standard; peptide; 23 AA.
XX
AC AAW83414;
XX
DT 26-FEB-1999 (first entry)
XX
DE Cell growth/adhesion promoting peptide #1.
XX
KW Cell growth; adhesion; promotion; medical treatment; injury;
KW biotissue; bone reinforcement; nerve regeneration; HMP resin.
XX
OS Synthetic.
XX
PN JP10316581-A.
XX
PD 02-DEC-1998.
XX
PF 15-MAY-1997; 97JP-0140885.
XX
PR 15-MAY-1997; 97JP-0140885.
XX
PA (KURS) KURARAY CO LTD.
XX
DR WPI; 1999-076400/07.
XX
PT Material for medical treatment comprises new peptide - used for
PT covering injuries, promoting adhesion of bio-tissues, bone
PT reinforcing and nerve regeneration
XX
PS Claim 1; Page 12; 14pp; Japanese.
XX
CC The present invention describes a material for medical treatment which
CC comprises one or more peptides of the formula XADEGJLMProQY, or their
CC salts, immobilised on a substrate: where X = H, CH3CO or CH3COLys;
CC A = Ser or Thr; D = Ile, Val or Leu; E = Lys or Arg; G = Ile, Val or
CC Leu; J = Gly or Ala; L = Ile, Val or Leu; M = Gly or Ala; Q = Gly, Ala
CC or Gly-Lys-Lys-Gly; Y = OH or NH2. Also described is an agent for cell
CC growth promotion and/or cell adhesion promotion containing the above
CC peptide or its salt as the active component. The peptide and its salt
CC can be used for covering injuries, promoting adhesion of biotissues,
CC bone reinforcing and nerve regeneration. The present sequence represents
CC a specifically claimed peptide of the present invention.
XX
SQ Sequence 23 AA;

  Query Match   100.0%; Score 131; DB 20; Length 23;
  Best Local Similarity 100.0%; Pred. No. 3.4e-08;
  Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db   1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 2
AAB12893
ID AAB12893 standard; peptide; 23 AA.
XX
AC AAB12893;
XX
DT 02-NOV-2000 (first entry)
XX
DE Nerve tissue regenerative peptide SEQ ID #8.
XX
KW Nerve regeneration; nerve cell proliferation; axon extension; treatment;
KW central nervous system disorder; peripheral nervous system disorder;
KW spinal disorder; head injury; cerebrovascular disorder.
XX
OS Synthetic.
XX
PN JP2000143531-A.
XX
PD 23-MAY-2000.
XX
PF 11-AUG-1999; 99JP-0227108.
XX
PR 09-SEP-1998; 98JP-0270498.
XX
PA (KURS) KURARAY CO LTD.
PA (NISH/) NISHIMURA Y.
PA (SUZU/) SUZUKI Y.
PA (TANI/) TANIHARA M.
XX
DR WPI; 2000-415772/36.
XX
PT New nerve regeneration material -
XX
PS Claim 2; Page 5; 17pp; Japanese.
XX
CC This invention relates to a new nerve regenerative material which
CC contains a peptide immobilised to a base which consists of a
CC polysaccharide gel such as alginic acid. Sequences AAB12886-B12899
CC represent examples of the peptides used in the nerve regeneration
CC material. The peptide containing material causes nerve cell
CC proliferation and also causes axonal extension. The material can be used
CC for the treatment of central or peripheral nervous system disorders,
CC spinal disorders, head injury or cerebrovascular disorders.
XX
SQ Sequence 23 AA;

  Query Match   100.0%; Score 131; DB 21; Length 23;
  Best Local Similarity 100.0%; Pred. No. 3.4e-08;
  Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db   1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 3
AAB70363
ID AAB70363 standard; peptide; 23 AA.

XX
AC AAB70363;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human thrombin receptor binding domain peptide SEQ ID NO:8.
XX
KW Neutrophil cell chemotactic; wound healing; inflammation; vulnerary;
KW antiinflammatory.
XX
OS Homo sapiens.
XX
PN US6184342-B1.
XX
PD 06-FEB-2001.
XX
PF 28-OCT-1994; 94US-0330594.
XX
PR 28-OCT-1994; 94US-0330594.
XX
PA (CHRY-) CHRYSALIS BIOTECHNOLOGY INC.
XX
PI Carney DH, Ramakrishnan S;
XX
DR WPI; 2001-202003/20.
XX
PT New synthetic neutrophil cell chemotactic peptides, useful for
PT generating antibodies for modulating neutrophil chemotaxis in immune
PT response and wound healing -
XX
PS Example 2; Column 6; 15pp; English.
XX
CC The present invention describes a synthetic peptide (I) which is a
CC neutrophil cell chemotactic agent. (I) has vulnerary and
CC antiinflammatory activities. (I) is useful as a potent neutrophil cell
CC chemotactic agent and for generating antibodies against the peptides,
CC which are useful for modulating neutrophil recruitment to a wound site
CC for enhancing or inhibiting inflammation and early effects of wound
CC healing. Neutrophil response to (I) is specific, since monocytes and
CC fibroblasts do not show any expression of the receptor to which (I)
CC binds. The present sequence represents a human thrombin receptor binding
CC domain peptide which is used in an example from the present invention.
XX
SQ Sequence 23 AA;

Query Match 100.0%; Score 131; DB 22; Length 23;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
||||||||||||||||||||||||
Db 1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 4
AAE22563
ID AAE22563 standard; peptide; 23 AA.
XX
AC AAE22563;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human thrombin high affinity receptor binding domain.
XX
KW Human; proteolytically activated receptor for thrombin; neutrophil;
KW chemotactic agent; PART; inflammation; wound healing; chemotaxis;
KW immune response; vulnerary; thrombin; receptor binding domain.
XX
OS Homo sapiens.
XX
PN US2002032314-A1.
XX
PD 14-MAR-2002.
XX
PF 05-FEB-2001; 2001US-0777328.
XX
PR 28-OCT-1994; 94US-0330594.
XX
PA (CHRY-) CHRYSALIS BIOTECHNOLOGY INC.
XX
PI Carney DH, Ramakrishnan S;
XX
DR WPI; 2002-371207/40.
XX
PT New synthetic peptide neutrophil cell chemotactic agents, useful for
PT stimulating or modulating neutrophil cell chemotactic migration,
PT particularly for modulating neutrophil recruitment during immune
PT response or in wound healing -
XX
PS Example 2; Page 3; 15pp; English.
XX
CC The present invention relates to novel synthetic peptides and antibodies
CC which are potent chemotactic agents for neutrophils. The peptides of the
CC invention mimic the activity and role of the cleavage fragment of the
CC proteolytically activated receptor for thrombin (PART). They are useful
CC for stimulating or modulating neutrophil cell chemotactic migration or
CC for generating an antibody. In particular, the peptides of the invention
CC are useful for modulating neutrophil recruitment to a wound site for
CC enhancing or inhibiting inflammation and early effects in wound healing.
CC They are also useful for modulated neutrophil chemotaxis in immune
CC response. The present sequence is high affinity receptor binding
CC domain of human thrombin. This peptide is used in the exemplification
CC of the invention.
XX
SQ Sequence 23 AA;

Query Match 100.0%; Score 131; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
||||||||||||||||||||||||
Db 1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 5
AAE20159
ID AAE20159 standard; peptide; 23 AA.
XX
AC AAE20159;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human thrombin peptide derivative #2.
XX
KW Cartilage growth; cartilage repair; arthritic joint; traumatic injury;
KW non-proteolytically activated thrombin receptor; NPAR; chondrocyte;
KW therapy; implantation; thrombin peptide; human.
XX
OS Homo sapiens.
XX
PN WO200207748-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US22668.
XX
PR 20-JUL-2000; 2000US-219800P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH, Crowther RS, Stiernberg J, Bergmann J;
XX
DR WPI; 2002-268953/31.
XX
PT Stimulating growth and repair of cartilage, useful for treating e.g.
PT arthritis, by local administration of an agonist of non-proteolytically
PT activated thrombin receptor -
XX
PS Claim 12; Page 25; 28pp; English.
XX
CC The invention relates to a method of stimulating growth and repair of
CC cartilage. The method involves administering to the site, an agonist
CC of non-proteolytically activated thrombin receptor (NPAR). The method
CC is used in human or veterinary medicine for the treatment of arthritic
CC joints and damage/loss of cartilage caused by traumatic injury. Also
CC chondrocytes may be cultured in presence of NPAR agonist to provide
CC cells for implantation at sites requiring growth/repair of cartilage.
CC The present sequence is human thrombin peptide derivative which serves
CC as a NPAR agonist.
XX
SQ Sequence 23 AA;

  Query Match   100.0%; Score 131; DB 23; Length 23;
  Best Local Similarity 100.0%; Pred. No. 3.4e-08;
  Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db   1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 6
AAM50858
ID AAM50858 standard; Peptide; 23 AA.
XX
AC AAM50858;
XX
DT 01-MAY-2002 (first entry)
XX
DE Thrombin-derived peptide used to promote cardiac tissue repair.
XX
KW Thrombin; revascularisation; vascular occlusion; tissue repair;
KW vulnerary; vasotropic; cardiant; angiogenesis; restenosis;
KW therapy; human.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Peptide   10..13
FT   /note= "thrombin receptor binding domain"
FT Peptide   12..23
FT   /note= "serine esterase conserved sequence"
XX
PN WO200204008-A2.
XX
PD 17-JAN-2002.
XX
PF 12-JUL-2001; 2001WO-US21944.
XX
PR 12-JUL-2000; 2000US-217583P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH;
XX
DR WPI; 2002-179665/23.
XX
PT Promoting cardiac tissue repair, stimulating revascularisation,
PT stimulating vascular endothelial cell proliferation, and inhibiting
PT vascular occlusion by using angiogenic thrombin derivative peptide -
XX
PS Claim 4; Page 19; 24pp; English.
XX
CC The present peptide comprises a thrombin-derived peptide, TP508,
CC that includes a thrombin receptor binding domain sequence (see also
CC AAM50856) and a serine esterase conserved sequence (see also
CC AAM50857). The peptide is used in a claimed method for promoting
CC cardiac tissue repair. It is administered during or following
CC cardiac surgery by injection into cardiac tissue, and may be
CC formulated as a sustained release formulation. The thrombin
CC derivative peptide is also used in claimed methods of stimulating
CC revascularisation, stimulating vascular endothelial cell
CC proliferation, inhibiting vascular occlusion, and inhibiting
CC restenosis following balloon angioplasty, in which case it may be
CC coated onto the catheter.
XX
SQ Sequence 23 AA;

```
  Query Match             100.0%;  Score 131;  DB 23;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-08;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db          1 AGYKPDEGKRGDACEGDSGGPFV 23


RESULT 7
AAW99115
ID   AAW99115 standard; protein; 116 AA.
XX
AC   AAW99115;
XX
DT   14-MAY-1999  (first entry)
XX
DE   Human zeta 2 prethrombin 2.
XX
KW   Prothrombin; exosite assay; anticoagulant; blood clot; thrombin;
KW   cardiovascular disease; stroke; haematological disorder.
XX
OS   Homo sapiens.
XX
PN   WO9855130-A1.
XX
PD   10-DEC-1998.
XX
PF   28-MAY-1998;   98WO-US10840.
XX
PR   08-APR-1998;   98US-0081030.
PR   06-JUN-1997;   97US-0048864.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Krishnaswamy S;
XX
DR   WPI; 1999-070237/06.
XX
PT   Exosite assay for agents that inhibit catalytic cleavage of
PT   prothrombin - at sites away from the active site of prothrombinase,
PT   also new inhibitors, potentially useful as anticoagulants
XX
PS   Disclosure; Page 44-45; 61pp; English.
XX
CC   An exosite assay has been developed for inhibition of the catalytic
CC   cleavage of prothrombin (PTh) to thrombin (Th) by prothrombinase (I), at
CC   a site remote from the catalytic site of (I) comprises: (a) preparing a
CC   solution containing 0.05-20 mu M substrate (S), that includes a protease
CC   cleavage site and exosite-binding determinant; 0.05-200 nM factor Va;
CC   30-500 micro M phospholipids (PL); test inhibitor (A) in buffer of pH
CC   7-9, containing 1-10 mM calcium ions but no calcium-chelating agent;
CC   (b) initiating catalytic cleavage of (S) by adding an aliquot of factor
CC   Xa (to final concentration 0.05-200 nM) so that there is an excess of Va
CC   over Xa, forming a S/(I) complex; (c) withdrawing aliquots of the
CC   reaction mixture, quenching them; and (d) assaying for concentration of
CC   Th. Alternatively, in the initial solution S is replaced by the same
CC   concentration of Xa (less than the amount of Va), and reaction is started
CC   by adding S. Also described in the present invention are inhibitors (A')
CC   having IC50 less than 1 mu M identified by this assay. (A') are
CC   potentially useful as a new class of anticoagulants for treatment of
CC   cardiovascular disease, stroke and haematological disorders. The method
CC   is based on the finding that exosite interactions are critical for
CC   substrate specificity in catalytic formation of Th. The present sequence
CC   represents human zeta 2 prethrombin 2.
XX
SQ   Sequence   116 AA;

  Query Match             100.0%;  Score 131;  DB 20;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db         45 AGYKPDEGKRGDACEGDSGGPFV 67


RESULT 8
AAW11545
ID   AAW11545 standard; Protein; 259 AA.
XX
AC   AAW11545;
XX
DT   01-OCT-1997  (first entry)
XX
DE   Human thrombin Asn99 mutant.
XX
KW   Prothrombin; mutant; mutein; platelet aggregation; blood clotting;
KW   coagulation; reduce; decrease; hirudin; heparin; anti-thrombin III;
KW   antagonist; D99N.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..259
FT                   /label= thrombin_Asn99
FT   Misc-difference 99
FT                   /note= "Wild-type Asp residue has been replaced by
FT                         Asn"
XX
PN   WO9641868-A2.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-AT00105.
XX
PR   13-JUN-1995;   95AT-0001006.
XX
PA   (IMMO ) IMMUNO AG.
XX
PI   Eibl J,  Falkner F,  Fischer B,  Mitterer A,  Schlokat U;
XX
```

DR WPI; 1997-065455/06.
XX
PT Prothrombin mutants with reduced clotting activity - useful as
PT antagonists of thrombin inhibitors or for anticoagulant therapy
XX
PS Example 3; Page -; 73pp; German.
XX
CC Prothrombin mutants having one or more changes in amino acid sequence
CC compared with the natural protein and having 0-10% (preferably 0-0.25%)
CC of the activity of the natural protein are claimed, provided that the
CC changes in amino acid sequence do not affect the capacity of the
CC mutants to bind to specific ligands and receptors. The mutants have
CC greatly reduced clotting activity and are useful as antagonists of
CC thrombin inhibitors such as hirudin, heparin and anti-thrombin III.
CC The mutations may also result in changes to the in vivo half-life
CC of prothrombin. The half-life may be reduced to less than 10 minutes
CC or the mutant prothrombin may have an extended half-life of more than
CC 1 hour, making it useful as an anticoagulant and to inhibit side-
CC effects of anti-coagulant treatment. They are converted to inactive
CC thrombin and are able to compete with native, active thrombin for
CC binding to receptors. The present sequence represents the thrombin
CC mutant which is derived by trypsin cleavage of a specifically
CC claimed human prothrombin mutant in which Asp at position 419 is
CC changed to Asn. The thrombin Asn99 mutant was found to have only
CC 0.24% of the activity of wild-type thrombin on a chromogenic
CC substrate.
CC (Note: This sequence does not appear in the specification and has
CC been produced by modifying the wild-type sequence of human
CC prothrombin which appears in figure 1).
XX
SQ Sequence 259 AA;

  Query Match   100.0%; Score 131; DB 18; Length 259;
  Best Local Similarity 100.0%; Pred. No. 2.9e-07;
  Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db   188 AGYKPDEGKRGDACEGDSGGPFV 210

RESULT 9
ABP60563
ID ABP60563 standard; protein; 259 AA.
XX
AC ABP60563;
XX
DT 28-MAR-2003 (first entry)
XX
DE Human thrombin variant W215A B-chain.
XX
KW Human; thrombin; W215A; anticoagulant; prothrombin; antithrombotic;
KW thrombus; protein C activation.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Misc-difference 229
FT   /note= "Wild-type Trp substituted by Ala"
XX
PN WO2002100337-A2.
XX
PD 19-DEC-2002.
XX
PF 07-JUN-2002; 2002WO-US18211.
XX
PR 08-JUN-2001; 2001US-297089P.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Gruber A, Hanson SR, Di Cera E;
XX
DR WPI; 2003-156907/15.
XX
PT New variant thrombin, useful as an antithrombotic agent for inhibiting
PT the formation of a thrombus, for determining the level of protein C
PT activation in a blood sample, or for determining the thrombogenic
PT potential of a patient -
XX
PS Claim 15; Fig 2; 95pp; English.
XX
CC The invention relates to a novel variant human thrombin. The thrombin
CC variant of the invention has anticoagulant activity. The variant thrombin
CC or prothrombin is useful as an antithrombotic agent for inhibiting the
CC formation of a thrombus. The variant thrombin is also useful for
CC determining the level of protein C activation in a blood sample or the
CC thrombogenic potential of a patient. The present sequence represents the
CC B-chain of the thrombin variant W215A.
XX
SQ Sequence 259 AA;

  Query Match   100.0%; Score 131; DB 24; Length 259;
  Best Local Similarity 100.0%; Pred. No. 2.9e-07;
  Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db   188 AGYKPDEGKRGDACEGDSGGPFV 210

RESULT 10
ABP60565
ID ABP60565 standard; protein; 259 AA.
XX
AC ABP60565;
XX
DT 28-MAR-2003 (first entry)
XX
DE Human thrombin variant W215A/E217A B-chain.
XX
KW Human; thrombin; W215A/E217A; anticoagulant; prothrombin; antithrombotic;
KW thrombus; protein C activation.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 227
FT                   /note= "Wild-type Trp substituted by Ala"
FT   Misc-difference 229
FT                   /note= "Wild-type Glu substituted by Ala"
XX
PN   WO2002100337-A2.
XX
PD   19-DEC-2002.
XX
PF   07-JUN-2002; 2002WO-US18211.
XX
PR   08-JUN-2001; 2001US-297089P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Gruber A,  Hanson SR,  Di Cera E;
XX
DR   WPI; 2003-156907/15.
DR   N-PSDB; ABZ25535.
XX
PT   New variant thrombin, useful as an antithrombotic agent for inhibiting
PT   the formation of a thrombus, for determining the level of protein C
PT   activation in a blood sample, or for determining the thrombogenic
PT   potential of a patient  -
XX
PS   Claim 2; Fig 4; 95pp; English.
XX
CC   The invention relates to a novel variant human thrombin. The thrombin
CC   variant of the invention has anticoagulant activity. The variant thrombin
CC   or prothrombin is useful as an antithrombotic agent for inhibiting the
CC   formation of a thrombus. The variant thrombin is also useful for
CC   determining the level of protein C activation in a blood sample or the
CC   thrombogenic potential of a patient. The present sequence represents the
CC   B-chain of the thrombin variant W215A/E217A (WE).
XX
SQ   Sequence   259 AA;

  Query Match           100.0%;  Score 131;  DB 24;  Length 259;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        188 AGYKPDEGKRGDACEGDSGGPFV 210


RESULT 11
AAR74775
ID   AAR74775 standard; Protein; 295 AA.
XX
AC   AAR74775;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Wild-type thrombin.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
DR   N-PSDB; AAQ92455.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities  -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Disclosure; Fig 1; 78pp; English.
XX
CC   The sequence reresents wild-type (reference) thrombin.  Mutants
CC   of this sequence (AAR74776-80 and AAR76033-41) have at least 80%
CC   homology with thrombin, and are capable of protein-C activation
CC   without significant fibrinogen clotting activity, and vice versa
CC   (specifically have a ratio of protein-C activity to fibrinogen
CC   clotting activity of less than 0.5 or greater than 2 compared to
CC   thrombin).  The mutant thrombin sequences, produced in recombinant
CC   cell culture or by in vitro methods, and are used to treat
CC   thrombotic conditions, particularly during cardiac bypass surgery
CC   and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;

  Query Match           100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity  100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246


RESULT 12
```

```
AAR74776
ID   AAR74776 standard; Protein; 295 AA.
XX
AC   AAR74776;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin K52A, R233A.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 88
FT                   /note= "Lys in wild-type"
FT   Misc-difference 269
FT                   /note= "Arg in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Claim 22; Page 63/3; 78pp; English.
XX
CC   The mutant thrombin sequence, generated by oligonucleotide-directed
CC   mutagenesis, has at least 80% homology with thrombin, and is
CC   capable of protein-C activation without significant fibrinogen
CC   clotting activity, and vice versa (specifically, it has a ratio
CC   of protein-C activity to fibrinogen clotting activity of less than
CC   0.5 or greater than 2 compared to thrombin).  The mutant thrombin
CC   is produced in recombinant cell culture or by in vitro methods,
CC   and is used to treat thrombotic conditions, particularly during
CC   cardiac bypass surgery and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;
```

```
  Query Match           100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity  100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246
```

```
RESULT 13
AAR74777
ID   AAR74777 standard; Protein; 295 AA.
XX
AC   AAR74777;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin E229D.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 265
FT                   /note= "Glu in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Claim 22; Page 63/3; 78pp; English.
XX
CC   The mutant thrombin sequence, generated by oligonucleotide-directed
CC   mutagenesis, has at least 80% homology with thrombin, and is
CC   capable of protein-C activation without significant fibrinogen
CC   clotting activity, and vice versa (specifically, it has a ratio
CC   of protein-C activity to fibrinogen clotting activity of less than
```

```
CC   0.5 or greater than 2 compared to thrombin).  The mutant thrombin
CC   is produced in recombinant cell culture or by in vitro methods,
CC   and is used to treat thrombotic conditions, particularly during
CC   cardiac bypass surgery and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;

  Query Match             100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246


RESULT 14
AAR74778
ID   AAR74778 standard; Protein; 295 AA.
XX
AC   AAR74778;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin E229F.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 265
FT                   /note= "Glu in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Claim 22; Page 63/3; 78pp; English.
XX
CC   The mutant thrombin sequence, generated by oligonucleotide-directed
CC   mutagenesis, has at least 80% homology with thrombin, and is
CC   capable of protein-C activation without significant fibrinogen
CC   clotting activity, and vice versa (specifically, it has a ratio
CC   of protein-C activity to fibrinogen clotting activity of less than
CC   0.5 or greater than 2 compared to thrombin).  The mutant thrombin
CC   is produced in recombinant cell culture or by in vitro methods,
CC   and is used to treat thrombotic conditions, particularly during
CC   cardiac bypass surgery and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;

  Query Match             100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246


RESULT 15
AAR74779
ID   AAR74779 standard; Protein; 295 AA.
XX
AC   AAR74779;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin E229S.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 265
FT                   /note= "Glu in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
```

XX
PI Gibbs CS, Leung LLK, Tsiang M;
XX
DR WPI; 1995-194103/25.
XX
PT Thrombin derivs with segregated pro- and anticoagulant activities -
PT useful for treating thrombotic disorders but also diagnosis,
PT treatment of tumours, etc.
XX
PS Claim 22; Page 63/3; 78pp; English.
XX
CC The mutant thrombin sequence, generated by oligonucleotide-directed
CC mutagenesis, has at least 80% homology with thrombin, and is
CC capable of protein-C activation without significant fibrinogen
CC clotting activity, and vice versa (specifically, it has a ratio
CC of protein-C activity to fibrinogen clotting activity of less than
CC 0.5 or greater than 2 compared to thrombin). The mutant thrombin
CC is produced in recombinant cell culture or by in vitro methods,
CC and is used to treat thrombotic conditions, particularly during
CC cardiac bypass surgery and in cases of septic shock.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 295 AA;

  Query Match   100.0%; Score 131; DB 16; Length 295;
  Best Local Similarity 100.0%; Pred. No. 3.3e-07;
  Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db   224 AGYKPDEGKRGDACEGDSGGPFV 246

Search completed: February 11, 2004, 14:53:25
Job time : 50.7097 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:49:07 ; Search time 15.5806 Seconds
                                    (without alignments)
                                    141.963 Million cell updates/sec

Title: US-10-050-611-3
Perfect score: 131
Sequence: 1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : PIR_76:*
            1: pir1:*
            2: pir2:*
            3: pir3:*
            4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	131	100.0	622	1	TBHU	thrombin (EC 3.4.2
2	127	96.9	236	2	C42696	thrombin (EC 3.4.2
3	124	94.7	625	1	TBBO	thrombin (EC 3.4.2
4	118	90.1	234	2	F42696	thrombin (EC 3.4.2
5	113	86.3	235	2	D42696	thrombin (EC 3.4.2
6	113	86.3	235	2	E42696	thrombin (EC 3.4.2
7	110	84.0	236	2	I42696	thrombin (EC 3.4.2
8	109	83.2	239	2	G42696	thrombin (EC 3.4.2
9	102	77.9	617	2	S10511	thrombin (EC 3.4.2
10	102	77.9	618	2	A35827	thrombin (EC 3.4.2
11	89	67.9	235	2	H42696	thrombin (EC 3.4.2
12	71.5	54.6	417	1	S00845	hepsin (EC 3.4.21.
13	71	54.2	461	1	KXHU	protein C (activat

14	70.5	53.8	482	1	EXRT	coagulation factor
15	70.5	53.8	638	1	KQHUP	plasma kallikrein
16	69.5	53.1	275	2	S40007	trypsin (EC 3.4.21
17	69.5	53.1	1524	2	T30337	polyprotein - Afri
18	68.5	52.3	161	2	I62744	coagulation factor
19	68.5	52.3	488	1	EXHU	coagulation factor
20	68.5	52.3	1019	2	A38738	coagulation factor
21	67.5	51.5	161	2	I48158	coagulation factor
22	67.5	51.5	282	2	I84621	coagulation factor
23	67.5	51.5	459	2	JQ0419	coagulation factor
24	67.5	51.5	475	1	EXCH	coagulation factor
25	67.5	51.5	638	1	KQMSPL	plasma kallikrein
26	67	51.1	225	2	S45356	probable serine pr
27	67	51.1	264	2	S32794	trypsin-like prote
28	66.5	50.8	309	2	B49878	coagulation factor
29	66.5	50.8	1004	2	T30338	oviductin (EC 3.4.
30	65.5	50.0	267	2	S40006	trypsin (EC 3.4.21
31	65.5	50.0	274	2	S35339	trypsin (EC 3.4.21
32	65.5	50.0	275	2	S40005	trypsin (EC 3.4.21
33	65.5	50.0	277	2	S35340	trypsin (EC 3.4.21
34	65.5	50.0	638	1	KQRTPL	plasma kallikrein
35	64.5	49.2	237	2	S55378	serine proteinase
36	64.5	49.2	238	1	TRWV5Y	trypsin-like prote
37	64	48.9	191	2	S54115	complement factor
38	64	48.9	246	1	DBHU	complement factor
39	64	48.9	456	1	KXBO	protein C (activat
40	64	48.9	2616	2	A57096	nudel protein prec
41	63.5	48.5	625	1	KFHU1	coagulation factor
42	63	48.1	461	1	JX0210	protein C (activat
43	62.5	47.7	375	1	A23689	limulus clotting e
44	62.5	47.7	416	1	S33777	hepsin (EC 3.4.21.
45	62.5	47.7	492	1	EXBO	coagulation factor

ALIGNMENTS

RESULT 1
TBHU
thrombin (EC 3.4.21.5) precursor [validated] - human
N;Alternate names: coagulation factor II
N;Contains: prothrombin
C;Species: Homo sapiens (man)
C;Date: 30-Nov-1980 #sequence_revision 22-Jul-1994 #text_change 08-Dec-2000
C;Accession: A29351; A00914; B00914; A37549; A37550; I51952
R;Degen, S.J.F.; Davie, E.W.
Biochemistry 26, 6165-6177, 1987
A;Title: Nucleotide sequence of the gene for human prothrombin.
A;Reference number: A29351; MUID:88077877; PMID:2825773
A;Accession: A29351
A;Molecule type: DNA
A;Residues: 1-622 <DEG>
A;Cross-references: GB:M17262; GB:M33691; NID:g558069; PIDN:AAC63054.1;
PID:g339641
R;Degen, S.J.F.; MacGillivray, R.T.A.; Davie, E.W.
Biochemistry 22, 2087-2097, 1983
A;Title: Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin.
A;Reference number: A00914; MUID:83231469; PMID:6305407
A;Accession: A00914
A;Molecule type: mRNA
A;Residues: 8-163,'N',165-622 <DE2>
A;Cross-references: GB:V00595; GB:J00307; NID:g37128; PIDN:CAA23842.1;
PID:g1335344
A;Accession: B00914
A;Molecule type: DNA
A;Residues: 188-311 <DE3>
R;Walz, D.A.; Hewett-Emmett, D.; Seegers, W.H.
Proc. Natl. Acad. Sci. U.S.A. 74, 1969-1972, 1977
A;Reference number: A37549; MUID:77193964; PMID:266717
A;Accession: A37549
A;Molecule type: protein
A;Residues: 44-118,'N',120,'S',122-163,'I',165-175,'A',177-182,'T',184-193,'MV',196-308,'EE',309-314 <WAL>
R;Butkowski, R.J.; Elion, J.; Downing, M.R.; Mann, K.G.
J. Biol. Chem. 252, 4942-4957, 1977
A;Title: Primary structure of human prethrombin 2 and alpha-thrombin.
A;Reference number: A37550; MUID:77207112; PMID:873923
A;Accession: A37550
A;Molecule type: protein
A;Residues: 315-334,'N',336-348,'N',350-368,'N',370-397,'N',399-413,'N',415-484,'N',486-493,'G',495-503,'Y',505-508,'S',510,'V',512-513,'D',515-528,'AL',531,'Q',533-622 <BUT>
R;Rabiet, M.J.; Blashill, A.; Furie, B.; Furie, B.C.
J. Biol. Chem. 261, 13210-13215, 1986
A;Reference number: A37551; MUID:87008532; PMID:3759958
A;Contents: annotation; activation cleavages
R;MacGillivray, R.T.; Irwin, D.M.; Guinto, E.R.; Stone, J.C.
Ann. N. Y. Acad. Sci. 485, 73-79, 1986
A;Title: Recombinant genetic approaches to functional mapping of thrombin.
A;Reference number: I51952; MUID:87182874; PMID:3471151
A;Accession: I51952
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2,'RI',5-100 <RES>
A;Cross-references: GB:M33031; NID:g190723; PIDN:AAA60220.1; PID:g190724
C;Comment: Thrombin, which cleaves bonds after Arg and Lys, converts fibrinogen to fibrin and activates factors V, VIII, XIII, and, in complex with thrombomodulin, protein C.
C;Comment: Prothrombin is activated on the surface of a phospholipid membrane that binds the amino end of prothrombin and factors Va and Xa in calcium-dependent interactions. The activation peptide(s) can be removed either by factor Xa or thrombin; the cleavage into light and heavy chains is by factor Xa. It is not known whether one or two smaller activation peptides, with additional cleavage after 314-Arg, are released in natural blood clotting.
C;Comment: The cleavage after Arg-198, observed in vitro, does not occur in plasma.
C;Comment: The gamma-carboxyglutamyl residues bind calcium ions, result from the carboxylation of glutamyl residues by microsomal vitamin K-dependent carboxylase, and are necessary for calcium-dependent interaction with the negatively charged phospholipid membrane surface.
C;Comment: The prothrombin precursor is synthesized in the liver.
C;Genetics:

A;Gene: GDB:F2
A;Cross-references: GDB:119894; OMIM:176930
A;Map position: 11p11-11q12
A;Introns: 27/1; 80/3; 89/1; 106/1; 141/2; 187/1; 292/1; 335/1; 377/2; 433/2; 491/2; 552/1; 575/3
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: acute phase; blood coagulation; calcium binding; carboxyglutamic acid; duplication; glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-43/Domain: propeptide #status predicted <PRO>
F;28-87/Domain: Gla domain homology <GLA>
F;44-622/Product: prothrombin #status experimental <MAT>
F;44-327/Domain: activation peptide #status experimental <APT>
F;108-186/Domain: kringle homology <KR1>
F;213-291/Domain: kringle homology <KR2>
F;328-363/Product: thrombin light chain #status experimental <LCH>
F;364-622/Product: thrombin heavy chain #status experimental <HCH>
F;364-613/Domain: trypsin homology <TRY>
F;49,50,57,59,62,63,68,69,72,75/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental
F;60-65,90-103,108-186,129-169,157-181,213-291,234-274,262-286/Disulfide bonds: #status predicted
F;121,143/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;336-482,536-550,564-594/Disulfide bonds: #status predicted
F;391-407/Disulfide bonds: #status experimental
F;406,462/Active site: His, Asp #status predicted
F;416/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;568/Active site: Ser #status experimental

  Query Match 100.0%; Score 131; DB 1; Length 622;
  Best Local Similarity 100.0%; Pred. No. 1.9e-10;
  Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
    |||||||||||||||||||||||
Db 551 AGYKPDEGKRGDACEGDSGGPFV 573

RESULT 2
C42696
thrombin (EC 3.4.21.5) B chain - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 17-Mar-1999
C;Accession: C42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: C42696
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-236 <BAN>
A;Cross-references: GB:M81396
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-227/Domain: trypsin homology (fragment) <TRY>

  Query Match 96.9%; Score 127; DB 2; Length 236;
  Best Local Similarity 95.7%; Pred. No. 2.6e-10;
  Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
    |||||:|||||||||||||||||
Db 165 AGYKPEEGKRGDACEGDSGGPFV 187

Search completed: February 11, 2004, 14:56:57
Job time : 16.5806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:36:52 ; Search time 9.64516 Seconds
(without alignments)
112.141 Million cell updates/sec

Title: US-10-050-611-3
Perfect score: 131
Sequence: 1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	131	100.0	622	1	THRB_HUMAN	P00734 homo sapien
2	124	94.7	625	1	THRB_BOVIN	P00735 bos taurus
3	102	77.9	617	1	THRB_RAT	P18292 rattus norv
4	102	77.9	618	1	THRB_MOUSE	P19221 mus musculu
5	73.5	56.1	290	1	MPN_HUMAN	Q9bqr3 homo sapien
6	71.5	54.6	417	1	HEPS_HUMAN	P05981 homo sapien
7	71.5	54.6	436	1	HEPS_MOUSE	O35453 mus musculu
8	71	54.2	161	1	PRTC_MACMU	Q28506 macaca mula
9	71	54.2	461	1	PRTC_HUMAN	P04070 homo sapien
10	70.5	53.8	638	1	KAL_HUMAN	P03952 homo sapien
11	70	53.4	281	1	TRYZ_DROER	P54630 drosophila
12	69.5	53.1	275	1	TRY3_ANOGA	P35037 anopheles g
13	68.5	52.3	488	1	FA10_HUMAN	P00742 homo sapien
14	68.5	52.3	1019	1	LFC_CARRO	Q26422 carcinoscor
15	68.5	52.3	1019	1	LFC_TACTR	P28175 tachypleus
16	68	51.9	458	1	PRTC_RABIT	Q28661 oryctolagus
17	67.5	51.5	282	1	FA9_RAT	P16296 rattus norv
18	67.5	51.5	459	1	FA9_MOUSE	P16294 mus musculu
19	67.5	51.5	475	1	FA10_CHICK	P25155 gallus gall
20	67.5	51.5	638	1	KAL_MOUSE	P26262 mus musculu
21	67	51.1	256	1	KLKF_HUMAN	Q9h2r5 homo sapien
22	67	51.1	264	1	VDP_BOMMO	Q07943 bombyx mori
23	66.5	50.8	455	1	TMS5_MOUSE	Q9er04 mus musculu
24	66.5	50.8	457	1	TMS5_HUMAN	Q9h3s3 homo sapien
25	65.5	50.0	267	1	TRY7_ANOGA	P35041 anopheles g
26	65.5	50.0	274	1	TRY1_ANOGA	P35035 anopheles g
27	65.5	50.0	275	1	TRY4_ANOGA	P35038 anopheles g
28	65.5	50.0	277	1	TRY2_ANOGA	P35036 anopheles g
29	65.5	50.0	638	1	KAL_RAT	P14272 rattus norv
30	65	49.6	157	1	PRTC_CANFA	Q28278 canis famil
31	65	49.6	157	1	PRTC_CAPHI	Q28315 capra hircu
32	65	49.6	157	1	PRTC_FELCA	Q28412 felis silve
33	65	49.6	157	1	PRTC_HORSE	Q28380 equus cabal
34	65	49.6	459	1	PRTC_PIG	Q9glp2 sus scrofa
35	64.5	49.2	238	1	TRY5_AEDAE	P29787 aedes aegyp
36	64.5	49.2	422	1	DES1_HUMAN	Q9ul52 homo sapien
37	64.5	49.2	490	1	FA10_RABIT	O19045 oryctolagus
38	64	48.9	253	1	CFAD_HUMAN	P00746 homo sapien
39	64	48.9	259	1	CFAD_PIG	P51779 sus scrofa
40	64	48.9	456	1	PRTC_BOVIN	P00745 bos taurus
41	64	48.9	875	1	NETR_HUMAN	P56730 homo sapien
42	64	48.9	2616	1	NDL_DROME	P98159 drosophila
43	63.5	48.5	625	1	FA11_HUMAN	P03951 homo sapien
44	63	48.1	256	1	TRYE_DROER	P54627 drosophila
45	63	48.1	461	1	PRTC_MOUSE	P33587 mus musculu

ALIGNMENTS

RESULT 1
THRB_HUMAN
ID THRB_HUMAN STANDARD; PRT; 622 AA.
AC P00734;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Prothrombin precursor (EC 3.4.21.5) (Coagulation factor II).
GN F2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88077877; PubMed=2825773;
RA Degen S.J.F., Davie E.W.;
RT "Nucleotide sequence of the gene for human prothrombin.";
RL Biochemistry 26:6165-6177(1987).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT MET-165.
RA Rieder M.J., Armel T.Z., Carrington D.P., Chung M.-W., Lee K.L.,
RA Ozuna M., Poel C.L., Toth E.J., Yi Q., Nickerson D.A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.

RN [3]
RP SEQUENCE OF 8-622 FROM N.A.
RX MEDLINE=83231469; PubMed=6305407;
RA Degen S.J.F., McGillivray R.T.A., Davie E.W.;
RT "Characterization of the complementary deoxyribonucleic acid and gene
RT coding for human prothrombin.";
RL Biochemistry 22:2087-2097(1983).
RN [4]
RP SEQUENCE OF 44-314.
RX MEDLINE=77193964; PubMed=266717;
RA Walz D.A., Hewett-Emmett D., Seegers W.H.;
RT "Amino acid sequence of human prothrombin fragments 1 and 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 74:1969-1972(1977).
RN [5]
RP SEQUENCE OF 315-622.
RX MEDLINE=77207112; PubMed=873923;
RA Butkowski R.J., Elion J., Downing M.R., Mann K.G.;
RT "Primary structure of human prethrombin 2 and alpha-thrombin.";
RL J. Biol. Chem. 252:4942-4957(1977).
RN [6]
RP PROCESSING.
RX MEDLINE=87008532; PubMed=3759958;
RA Rabiet M.J., Blashill A., Furie B., Furie B.C.;
RT "Prothrombin fragment 1 X 2 X 3, a major product of prothrombin
RT activation in human plasma.";
RL J. Biol. Chem. 261:13210-13215(1986).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=90059942; PubMed=2583108;
RA Bode W., Mayr I., Baumann U., Huber R., Stone S.R., Hofsteenge J.;
RT "The refined 1.9 A crystal structure of human alpha-thrombin:
RT interaction with D-Phe-Pro-Arg chloromethylketone and significance of
RT the Tyr-Pro-Pro-Trp insertion segment.";
RL EMBO J. 8:3467-3475(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=90327074; PubMed=2374926;
RA Rydel T.J., Ravichandran K.G., Tulinsky A., Bode W., Huber R.,
RA Roitsch C., Fenton J.W. II;
RT "The structure of a complex of recombinant hirudin and human alpha-
RT thrombin.";
RL Science 249:277-280(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=94350942; PubMed=8071320;
RA Rydel T.J., Yin M., Padmanabhan K.P., Blankenship D.T., Cardin A.D.,
RA Correa P.E., Fenton J.W. II, Tulinsky A.;
RT "Crystallographic structure of human gamma-thrombin.";
RL J. Biol. Chem. 269:22000-22006(1994).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=97357286; PubMed=9214615;
RA van de Locht A., Bode W., Huber R., le Bonniec B.F., Stone S.R.,
RA Esmon C.T., Stubbs M.T.;
RT "The thrombin E192Q-BPTI complex reveals gross structural
RT rearrangements: implications for the interaction with antithrombin
RT and thrombomodulin.";
RL EMBO J. 16:2977-2984(1997).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 328-601.
RX MEDLINE=99162521; PubMed=10051558;
RA Guinto E.R., Caccia S., Rose T., Fuetterer K., Waksman G., di Cera E.;
RT "Unexpected crucial role of residue 225 in serine proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1852-1857(1999).
RN [12]
RP VARIANT BARCELONA.
RX MEDLINE=87033739; PubMed=3771562;
RA Rabiet M.-J., Furie B.C., Furie B.;
RT "Molecular defect of prothrombin Barcelona. Substitution of cysteine
RT for arginine at residue 273.";
RL J. Biol. Chem. 261:15045-15048(1986).
RN [13]
RP VARIANT FRANKFURT.
RX MEDLINE=95313001; PubMed=7792730;
RA Degen S.J.F., McDowell S.A., Sparks L.M., Scharrer I.;
RT "Prothrombin Frankfurt: a dysfunctional prothrombin characterized by
RT substitution of Glu-466 by Ala.";
RL Thromb. Haemost. 73:203-209(1995).
RN [14]
RP VARIANTS HIMI-1 AND HIMI-2.
RX MEDLINE=93043342; PubMed=1421398;
RA Morishita E., Saito M., Kumabashiri I., Asakura H., Matsuda T.,
RA Yamaguchi K.;
RT "Prothrombin Himi: a compound heterozygote for two dysfunctional
RT prothrombin molecules (Met-337-->Thr and Arg-388-->His).";
RL Blood 80:2275-2280(1992).
RN [15]
RP VARIANT PADUA-1.
RX MEDLINE=95169898; PubMed=7865694;
RA James H.L., Kim D.J., Zheng D.-Q., Girolami A.;
RT "Prothrombin Padua I: incomplete activation due to an amino acid
RT substitution at a factor Xa cleavage site.";
RL Blood Coagul. Fibrinolysis 5:841-844(1994).
RN [16]
RP VARIANT QUICK-1.
RX MEDLINE=89207504; PubMed=3242619;
RA Henriksen R.A., Mann K.G.;
RT "Identification of the primary structural defect in the dysthrombin
RT thrombin Quick I: substitution of cysteine for arginine-382.";
RL Biochemistry 27:9160-9165(1988).
RN [17]
RP VARIANT QUICK-2.
RX MEDLINE=89247398; PubMed=2719946;
RA Henriksen R.A., Mann K.G.;
RT "Substitution of valine for glycine-558 in the congenital dysthrombin
RT thrombin Quick II alters primary substrate specificity.";
RL Biochemistry 28:2078-2082(1989).
RN [18]
RP VARIANT SALAKTA.
RX MEDLINE=92378975; PubMed=1354985;
RA Miyata T., Aruga R., Umeyama H., Bezeaud A., Guillin M.-C.,
RA Iwanaga S.;
RT "Prothrombin Salakta: substitution of glutamic acid-466 by alanine
RT reduces the fibrinogen clotting activity and the esterase activity.";

```
RL   Biochemistry 31:7457-7462(1992).
RN   [19]
RP   VARIANT TOKUSHIMA.
RX   MEDLINE=87185407; PubMed=3567158;
RA   Miyata T., Morita T., Inomoto T., Kawauchi S., Shirakami A.,
RA   Iwanaga S.;
RT   "Prothrombin Tokushima, a replacement of arginine-418 by tryptophan
RT   that impairs the fibrinogen clotting activity of derived thrombin
RT   Tokushima.";
RL   Biochemistry 26:1117-1122(1987).
RN   [20]
RP   VARIANT TOKUSHIMA.
RX   MEDLINE=87101511; PubMed=3801671;
RA   Inomoto T., Shirakami A., Kawauchi S., Shigekiyo T., Saito S.,
RA   Miyoshi K., Morita T., Iwanaga S.;
RT   "Prothrombin Tokushima: characterization of dysfunctional thrombin
RT   derived from a variant of human prothrombin.";
RL   Blood 69:565-569(1987).
RN   [21]
RP   VARIANT TOKUSHIMA.
RX   MEDLINE=92256895; PubMed=1349838;
RA   Iwahana H., Yoshimoto K., Shigekiyo T., Shirakami A., Saito S.,
RA   Itakura M.;
RT   "Detection of a single base substitution of the gene for prothrombin
RT   Tokushima. The application of PCR-SSCP for the genetic and molecular
RT   analysis of dysprothrombinemia.";
RL   Int. J. Hematol. 55:93-100(1992).
RN   [22]
RP   VARIANT TYPE-3.
RX   MEDLINE=83204687; PubMed=6405779;
RA   Board P.G., Shaw D.C.;
RT   "Determination of the amino acid substitution in human prothrombin
RT   type 3 (157 Glu leads to Lys) and the localization of a third
RT   thrombin cleavage site.";
RL   Br. J. Haematol. 54:245-254(1983).
RN   [23]
RP   VARIANTS MET-165 AND THR-386.
RX   MEDLINE=99318093; PubMed=10391209;
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RT   "Characterization of single-nucleotide polymorphisms in coding regions
RT   of human genes.";
RL   Nat. Genet. 22:231-238(1999).
RN   [24]
RP   ERRATUM.
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RL   Nat. Genet. 23:373-373(1999).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Gly; activates
CC       fibrinogen to fibrin and releases fibrinopeptide A and B.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION
```

```
  Query Match            100.0%;  Score 131;  DB 1;  Length 622;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        551 AGYKPDEGKRGDACEGDSGGPFV 573
```

```
RESULT 2
THRB_BOVIN
ID   THRB_BOVIN     STANDARD;      PRT;   625 AA.
AC   P00735;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Prothrombin precursor (EC 3.4.21.5).
GN   F2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88245190; PubMed=3379642;
RA   Irwin D.M., Robertson K.A., Macgillivray R.T.A.;
RT   "Structure and evolution of the bovine prothrombin gene.";
RL   J. Mol. Biol. 200:31-45(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84203525; PubMed=6326805;
RA   McGillivray R.T.A., Davie E.W.;
RT   "Characterization of bovine prothrombin mRNA and its translation
RT   product.";
RL   Biochemistry 23:1626-1634(1984).
RN   [3]
RP   SEQUENCE OF 44-625, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RA   Magnusson S., Sottrup-Jensen L., Petersen T.E., Claeys H.;
RL   (In) Hemker H.C., Veltkamp J.J. (eds.);
RL   Boerhaave symposium on prothrombin and related coagulation factors,
RL   pp.25-46, Leiden University Press, Leiden (1975).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE=86296631; PubMed=3741841;
RA   Park C.H., Tulinsky A.;
RT   "Three-dimensional structure of the kringle sequence: structure of
RT   prothrombin fragment 1.";
RL   Biochemistry 25:3977-3982(1986).
```

```
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE=91311686; PubMed=1856869;
RA   Seshadri T.-P., Tulinsky A., Skrzypczak-Jankun E., Park C.H.;
RT   "Structure of bovine prothrombin fragment 1 refined at 2.25-A
RT   resolution.";
RL   J. Mol. Biol. 220:481-494(1991).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF ACTIVATION PEPTIDE 1.
RX   MEDLINE=92190185; PubMed=1547238;
RA   Soriano-Garcia M., Padmanabhan K., de Vos A.M., Tulinsky A.;
RT   "The Ca2+ ion and membrane binding structure of the Gla domain of Ca-
RT   prothrombin fragment 1.";
RL   Biochemistry 31:2554-2566(1992).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=92218459; PubMed=1560020;
RA   Martin P.D., Robertson W., Turk D., Huber R., Bode W., Edwards B.F.P.;
RT   "The structure of residues 7-16 of the A alpha-chain of human
RT   fibrinogen bound to bovine thrombin at 2.3-A resolution.";
RL   J. Biol. Chem. 267:7911-7920(1992).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=92389319; PubMed=1518046;
RA   Brandstetter H., Turk D., Hoeffken H.W., Grosse D., Stuerzebecher J.,
RA   Martin P.D., Edwards B.F.P., Bode W.;
RT   "Refined 2.3 A X-ray crystal structure of bovine thrombin complexes
RT   formed with the benzamidine and arginine-based thrombin inhibitors
RT   NAPAP, 4-TAPAP and MQPA. A starting point for improving
RT   antithrombotics.";
RL   J. Mol. Biol. 226:1085-1089(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF COMPLEX WITH ORNITHODORIN.
RX   MEDLINE=97102783; PubMed=8947023;
RA   van de Locht A., Stubbs M.T., Bode W., Friedrich T., Bollschweiler C.,
RA   Hoffken W., Huber R.;
RT   "The ornithodorin-thrombin crystal structure, a key to the TAP
RT   enigma?";
RL   EMBO J. 15:6011-6017(1996).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH TRIABIN.
RX   MEDLINE=98004486; PubMed=9342325;
RA   Fuentes-Prior P., Noeske-Jungblut C., Donner P., Schleuning W.D.,
RA   Huber R., Bode W.;
RT   "Structure of the thrombin complex with triabin, a lipocalin-like
RT   exosite-binding inhibitor derived from a triatomine bug.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11845-11850(1997).
RN   [11]
RP   GENE STRUCTURE.
RX   MEDLINE=86077733; PubMed=3000440;
RA   Irwin D.M., Ahern K.G., Pearson G.D., McGillivray R.T.A.;
RT   "Characterization of the bovine prothrombin gene.";
RL   Biochemistry 24:6854-6861(1985).
CC   -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC       FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC       AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Gly; activates
CC       fibrinogen to fibrin and releases fibrinopeptide A and B.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC   -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC       RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC       ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC       ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC       CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION
CC       OF PROTHROMBIN TO THROMBIN.
CC   -!- MISCELLANEOUS: PROTHROMBIN IS ACTIVATED ON THE SURFACE OF A
CC       PHOSPHOLIPID MEMBRANE THAT BINDS THE AMINO END OF PROTHROMBIN &
CC       FACTORS VA & XA IN CA-DEPENDENT INTERACTIONS; FACTOR XA REMOVES
CC       THE ACTIVATION PEPTIDE & CLEAVES THE REMAINING PART INTO LIGHT &
CC       HEAVY CHAINS. THE ACTIVATION PROCESS STARTS SLOWLY BECAUSE FACTOR
CC       V ITSELF HAS TO BE ACTIVATED BY THE INITIAL, SMALL AMOUNTS OF
CC       THROMBIN.
CC   -!- MISCELLANEOUS: THROMBIN CAN ITSELF CLEAVE THE AMINO TERMINAL
CC       FRAGMENT (FRAGMENT 1) OF THE PROTHROMBIN, PRIOR TO ITS ACTIVATION
CC       BY FACTOR XA.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   -!- SIMILARITY: Contains 2 kringle domains.
CC   -!- DATABASE: NAME=ProZyme technical fact sheet;
CC       WWW="http://www.prozyme.com/technical/thrombindata.html".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; V00135; CAA23451.1; -.
DR   EMBL; J00041; AAA30781.1; -.
DR   PIR; S02537; TBBO.
DR   PDB; 1BBR; 31-JAN-94.
DR   PDB; 1ETR; 31-JAN-94.
DR   PDB; 1ETS; 31-JAN-94.
DR   PDB; 1ETT; 31-JAN-94.
DR   PDB; 1HRT; 31-JAN-94.
DR   PDB; 2PF1; 31-JAN-94.
DR   PDB; 2PF2; 31-JAN-94.
DR   PDB; 2SPT; 31-MAY-94.
DR   PDB; 1MKW; 07-JUL-97.
DR   PDB; 1MKX; 07-JUL-97.
DR   PDB; 1TBQ; 14-OCT-96.
DR   PDB; 1TBR; 14-OCT-96.
DR   PDB; 1TOC; 23-JUL-97.
DR   PDB; 1VIT; 21-APR-97.
DR   PDB; 1YCP; 06-MAY-98.
DR   PDB; 1A0H; 17-JUN-98.
DR   PDB; 1AVG; 16-FEB-99.
DR   PDB; 1BTH; 24-DEC-97.
DR   PDB; 1ID5; 12-SEP-01.
DR   PDB; 1UVT; 19-NOV-97.
DR   PDB; 2HPP; 31-JAN-94.
DR   MEROPS; S01.217; -.
```

```
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003966; Prothrombin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00051; kringle; 2.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00001; GLABLOOD.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR01505; PROTHROMBIN.
DR   ProDom; PD000395; Kringle; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00130; KR; 2.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00021; KRINGLE_1; 2.
DR   PROSITE; PS50070; KRINGLE_2; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Blood coagulation; Plasma; Calcium-binding; Glycoprotein; Repeat;
KW   Vitamin K; Zymogen; Gamma-carboxyglutamic acid; Acute phase; Liver;
KW   Hydrolase; Serine protease; Kringle; Signal; 3D-structure.
FT   SIGNAL       1     24       POTENTIAL.
FT   PROPEP      25     43
FT   CHAIN       44    625       PROTHROMBIN.
FT   PEPTIDE     44    199       ACTIVATION PEPTIDE (FRAGMENT 1).
FT   PEPTIDE    200    317       ACTIVATION PEPTIDE (FRAGMENT 2).
FT   CHAIN      318    366       THROMBIN LIGHT CHAIN (A).
FT   CHAIN      367    625       THROMBIN HEAVY CHAIN (B).
FT   DOMAIN     109    187       KRINGLE 1.
FT   DOMAIN     214    292       KRINGLE 2.
FT   DOMAIN     367    625       SERINE PROTEASE.
FT   SITE       199    200       CLEAVAGE (BY THROMBIN).
FT   SITE       317    318       CLEAVAGE (BY FACTOR XA).
FT   SITE       366    367       CLEAVAGE (BY FACTOR XA).
FT   ACT_SITE   409    409       CHARGE RELAY SYSTEM.
FT   ACT_SITE   465    465       CHARGE RELAY SYSTEM.
FT   ACT_SITE   571    571       CHARGE RELAY SYSTEM.
FT   MOD_RES     50     50       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     51     51       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     58     58       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     60     60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     63     63       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     64     64       GAMMA-CARBOXYGLUTAMIC ACID.

  Query Match             94.7%;  Score 124;  DB 1;  Length 625;
  Best Local Similarity   95.7%;  Pred. No. 1.9e-09;
  Matches   22;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              ||||| ||||||||||||||||||
Db        554 AGYKPGEGKRGDACEGDSGGPFV 576
```

Search completed: February 11, 2004, 14:54:04
Job time : 10.6452 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:47:57 ; Search time 39.3226 Seconds
(without alignments)
150.936 Million cell updates/sec

Title: US-10-050-611-3
Perfect score: 131
Sequence: 1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	96.9	235	6	Q28731	Q28731 oryctolagus
2	118	90.1	235	13	Q90387	Q90387 cynops pyrr
3	113	86.3	235	13	Q91004	Q91004 gecko gecko
4	113	86.3	607	13	Q91001	Q91001 gallus gall
5	113	86.3	608	13	Q9PTW7	Q9ptw7 struthio ca
6	109	83.2	239	13	Q91218	Q91218 oncorhynchu
7	105	80.2	420	13	Q90504	Q90504 eptatretus
8	98	74.8	172	13	Q9DFD1	Q9dfd1 oncorhynchu
9	92	70.2	234	13	Q90244	Q90244 acipenser t
10	72.5	55.3	389	13	Q9PVX7	Q9pvx7 xenopus lae
11	72.5	55.3	974	13	Q90WD8	Q90wd8 bufo japoni
12	71.5	54.6	435	11	Q9CW97	Q9cw97 mus musculu
13	71.5	54.6	799	11	Q9DBI0	Q9dbi0 mus musculu
14	71.5	54.6	802	4	Q8IUE2	Q8iue2 homo sapien
15	71.5	54.6	811	4	Q8IU80	Q8iu80 homo sapien
16	71	54.2	195	4	Q8J008	Q8j008 homo sapien
17	71	54.2	195	4	Q8J007	Q8j007 homo sapien
18	71	54.2	195	4	Q8J006	Q8j006 homo sapien
19	71	54.2	195	4	Q8IXB4	Q8ixb4 homo sapien
20	71	54.2	211	4	Q8J009	Q8j009 homo sapien
21	70.5	53.8	161	11	Q63109	Q63109 rattus norv
22	70.5	53.8	259	5	Q9XY61	Q9xy61 ctenocephal
23	70.5	53.8	267	5	Q9BK47	Q9bk47 luidia foli
24	70.5	53.8	481	11	O54740	O54740 mus musculu
25	70.5	53.8	481	11	Q99L32	Q99132 mus musculu
26	70.5	53.8	481	11	O88947	O88947 mus musculu
27	70.5	53.8	482	11	Q63207	Q63207 rattus norv
28	70	53.4	378	5	Q8SY50	Q8sy50 drosophila
29	69.5	53.1	200	11	Q924U6	Q924u6 mus musculu
30	69.5	53.1	1524	13	Q91674	Q91674 xenopus lae
31	68.5	52.3	161	6	Q28511	Q28511 macaca mula
32	68.5	52.3	236	5	Q9TYH3	Q9tyh3 schistosoma
33	68.5	52.3	488	5	Q9TYH4	Q9tyh4 schistosoma
34	68.5	52.3	766	4	Q8NBY4	Q8nby4 homo sapien
35	68.5	52.3	1019	5	Q8T9S1	Q8t9s1 tachypleus
36	68.5	52.3	1083	5	Q26423	Q26423 carcinoscor
37	68	51.9	686	13	Q9DGC2	Q9dgc2 cyprinus ca
38	67.5	51.5	156	5	O16007	O16007 schistosoma
39	67.5	51.5	161	11	Q60546	Q60546 mesocricetu
40	67.5	51.5	264	5	O02569	O02569 culex quinq
41	67.5	51.5	328	11	Q8BJR6	Q8bjr6 mus musculu
42	67.5	51.5	370	5	Q9VA44	Q9va44 drosophila
43	67.5	51.5	387	5	Q9XY57	Q9xy57 ctenocephal
44	67.5	51.5	474	13	Q8JHC8	Q8jhc8 brachydanio
45	67.5	51.5	638	11	Q8R0P5	Q8r0p5 mus musculu

Search completed: February 11, 2004, 14:56:05
Job time : 40.3226 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:35:52 ; Search time 49.7097 Seconds
(without alignments)
73.441 Million cell updates/sec

Title: US-10-050-611-4
Perfect score: 131
Sequence: 1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	131	100.0	23	20	AAW83414	Cell growth/adhesi
2	131	100.0	23	21	AAB12893	Nerve tissue regen
3	131	100.0	23	22	AAB70363	Human thrombin rec
4	131	100.0	23	23	AAE22563	Human thrombin hig
5	131	100.0	23	23	AAE20159	Human thrombin pep
6	131	100.0	23	23	AAM50858	Thrombin-derived p
7	131	100.0	116	20	AAW99115	Human zeta 2 preth
8	131	100.0	259	18	AAW11545	Human thrombin Asn
9	131	100.0	259	24	ABP60563	Human thrombin var
10	131	100.0	259	24	ABP60565	Human thrombin var
11	131	100.0	295	16	AAR74775	Wild-type thrombin
12	131	100.0	295	16	AAR74776	Mutant thrombin K5
13	131	100.0	295	16	AAR74777	Mutant thrombin E2
14	131	100.0	295	16	AAR74778	Mutant thrombin E2
15	131	100.0	295	16	AAR74779	Mutant thrombin E2
16	131	100.0	295	16	AAR74780	Mutant thrombin E2
17	131	100.0	295	16	AAR76033	Mutant thrombin E2
18	131	100.0	295	16	AAR76034	Mutant thrombin R2
19	131	100.0	295	16	AAR76035	Mutant thrombin R2
20	131	100.0	295	16	AAR76036	Mutant thrombin R2
21	131	100.0	295	16	AAR76037	Mutant thrombin W5
22	131	100.0	295	16	AAR76038	Mutant thrombin K5
23	131	100.0	295	16	AAR76039	Mutant thrombin W5
24	131	100.0	295	16	AAR76040	Mutant thrombin W5
25	131	100.0	295	18	AAW22892	Human mature throm
26	131	100.0	295	21	AAB08633	Amino acid sequenc
27	131	100.0	295	24	ABP60562	Human thrombin var
28	131	100.0	295	24	ABP60564	Human thrombin var
29	131	100.0	308	20	AAW99109	Human prethrombin
30	131	100.0	376	14	AAR41797	CD4/Thrombin fusio
31	131	100.0	376	20	AAY42789	Human CD4/thrombin
32	131	100.0	376	23	AAU10703	Human CD4-thrombin
33	131	100.0	579	14	AAR35763	Prothrombin (PT).
34	131	100.0	579	18	AAW11546	Human prothrombin
35	131	100.0	579	18	AAW11544	Human prothrombin
36	131	100.0	579	20	AAW99108	Human prothrombin.
37	131	100.0	615	14	AAR38741	Human prothrombin.
38	131	100.0	615	17	AAR96216	Human prothrombin.
39	131	100.0	615	17	AAR90377	Human prothrombin.
40	131	100.0	622	18	AAW11543	Human preprothromb
41	131	100.0	622	20	AAY49566	Platelet membrane
42	131	100.0	622	24	ABG74671	Human F2 protein.
43	124	94.7	111	20	AAW99113	Bovine zeta 2 pret
44	124	94.7	308	20	AAW99107	Bovine prethrombin
45	124	94.7	582	20	AAW99106	Bovine prothrombin

ALIGNMENTS

```
RESULT 1
AAW83414
ID   AAW83414 standard; peptide; 23 AA.
XX
AC   AAW83414;
XX
DT   26-FEB-1999  (first entry)
XX
DE   Cell growth/adhesion promoting peptide #1.
XX
KW   Cell growth; adhesion; promotion; medical treatment; injury;
KW   biotissue; bone reinforcement; nerve regeneration; HMP resin.
XX
OS   Synthetic.
XX
PN   JP10316581-A.
XX
PD   02-DEC-1998.
XX
PF   15-MAY-1997;   97JP-0140885.
XX
PR   15-MAY-1997;   97JP-0140885.
XX
PA   (KURS ) KURARAY CO LTD.
XX
DR   WPI; 1999-076400/07.
XX
PT   Material for medical treatment comprises new peptide - used for
PT   covering injuries, promoting adhesion of bio-tissues, bone
PT   reinforcing and nerve regeneration
XX
PS   Claim 1; Page 12; 14pp; Japanese.
XX
CC   The present invention describes a material for medical treatment which
CC   comprises one or more peptides of the formula XADEGJLMProQY, or their
CC   salts, immobilised on a substrate: where X = H, CH3CO or CH3COLys;
CC   A = Ser or Thr; D = Ile, Val or Leu; E = Lys or Arg; G = Ile, Val or
CC   Leu; J = Gly or Ala; L = Ile, Val or Leu; M = Gly or Ala; Q = Gly, Ala
CC   or Gly-Lys-Lys-Gly; Y = OH or NH2. Also described is an agent for cell
CC   growth promotion and/or cell adhesion promotion containing the above
CC   peptide or its salt as the active component. The peptide and its salt
CC   can be used for covering injuries, promoting adhesion of biotissues,
CC   bone reinforcing and nerve regeneration. The present sequence represents
CC   a specifically claimed peptide of the present invention.
XX
SQ   Sequence   23 AA;

  Query Match            100.0%;  Score 131;  DB 20;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-08;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db          1 AGYKPDEGKRGDACEGDSGGPFV 23


RESULT 2
```

```
AAB12893
ID   AAB12893 standard; peptide; 23 AA.
XX
AC   AAB12893;
XX
DT   02-NOV-2000  (first entry)
XX
DE   Nerve tissue regenerative peptide SEQ ID #8.
XX
KW   Nerve regeneration; nerve cell proliferation; axon extension; treatment;
KW   central nervous system disorder; peripheral nervous system disorder;
KW   spinal disorder; head injury; cerebrovascular disorder.
XX
OS   Synthetic.
XX
PN   JP2000143531-A.
XX
PD   23-MAY-2000.
XX
PF   11-AUG-1999;   99JP-0227108.
XX
PR   09-SEP-1998;   98JP-0270498.
XX
PA   (KURS ) KURARAY CO LTD.
PA   (NISH/) NISHIMURA Y.
PA   (SUZU/) SUZUKI Y.
PA   (TANI/) TANIHARA M.
XX
DR   WPI; 2000-415772/36.
XX
PT   New nerve regeneration material  -
XX
PS   Claim 2; Page 5; 17pp; Japanese.
XX
CC   This invention relates to a new nerve regenerative material which
CC   contains a peptide immobilised to a base which consists of a
CC   polysaccharide gel such as alginic acid. Sequences AAB12886-B12899
CC   represent examples of the peptides used in the nerve regeneration
CC   material. The peptide containing material causes nerve cell
CC   proliferation and also causes axonal extension. The material can be used
CC   for the treatment of central or peripheral nervous system disorders,
CC   spinal disorders, head injury or cerebrovascular disorders.
XX
SQ   Sequence   23 AA;

  Query Match            100.0%;  Score 131;  DB 21;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-08;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db          1 AGYKPDEGKRGDACEGDSGGPFV 23


RESULT 3
AAB70363
ID   AAB70363 standard; peptide; 23 AA.
```

XX

AC AAB70363;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human thrombin receptor binding domain peptide SEQ ID NO:8.

XX

KW Neutrophil cell chemotactic; wound healing; inflammation; vulnerary;
KW antiinflammatory.

XX

OS Homo sapiens.

XX

PN US6184342-B1.

XX

PD 06-FEB-2001.

XX

PF 28-OCT-1994; 94US-0330594.

XX

PR 28-OCT-1994; 94US-0330594.

XX

PA (CHRY-) CHRYSALIS BIOTECHNOLOGY INC.

XX

PI Carney DH, Ramakrishnan S;

XX

DR WPI; 2001-202003/20.

XX

PT New synthetic neutrophil cell chemotactic peptides, useful for
PT generating antibodies for modulating neutrophil chemotaxis in immune
PT response and wound healing -

XX

PS Example 2; Column 6; 15pp; English.

XX

CC The present invention describes a synthetic peptide (I) which is a
CC neutrophil cell chemotactic agent. (I) has vulnerary and
CC antiinflammatory activities. (I) is useful as a potent neutrophil cell
CC chemotactic agent and for generating antibodies against the peptides,
CC which are useful for modulating neutrophil recruitment to a wound site
CC for enhancing or inhibiting inflammation and early effects of wound
CC healing. Neutrophil response to (I) is specific, since monocytes and
CC fibroblasts do not show any expression of the receptor to which (I)
CC binds. The present sequence represents a human thrombin receptor binding
CC domain peptide which is used in an example from the present invention.

XX

SQ Sequence 23 AA;

Query Match 100.0%; Score 131; DB 22; Length 23;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
||||||||||||||||||||||||
Db 1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 4
AAE22563
ID AAE22563 standard; peptide; 23 AA.

XX

AC AAE22563;

XX

DT 26-JUL-2002 (first entry)

XX

DE Human thrombin high affinity receptor binding domain.

XX

KW Human; proteolytically activated receptor for thrombin; neutrophil;
KW chemotactic agent; PART; inflammation; wound healing; chemotaxis;
KW immune response; vulnerary; thrombin; receptor binding domain.

XX

OS Homo sapiens.

XX

PN US2002032314-A1.

XX

PD 14-MAR-2002.

XX

PF 05-FEB-2001; 2001US-0777328.

XX

PR 28-OCT-1994; 94US-0330594.

XX

PA (CHRY-) CHRYSALIS BIOTECHNOLOGY INC.

XX

PI Carney DH, Ramakrishnan S;

XX

DR WPI; 2002-371207/40.

XX

PT New synthetic peptide neutrophil cell chemotactic agents, useful for
PT stimulating or modulating neutrophil cell chemotactic migration,
PT particularly for modulating neutrophil recruitment during immune
PT response or in wound healing -

XX

PS Example 2; Page 3; 15pp; English.

XX

CC The present invention relates to novel synthetic peptides and antibodies
CC which are potent chemotactic agents for neutrophils. The peptides of the
CC invention mimic the activity and role of the cleavage fragment of the
CC proteolytically activated receptor for thrombin (PART). They are useful
CC for stimulating or modulating neutrophil cell chemotactic migration or
CC for generating an antibody. In particular, the peptides of the invention
CC are useful for modulating neutrophil recruitment to a wound site for
CC enhancing or inhibiting inflammation and early effects in wound healing.
CC They are also useful for modulated neutrophil chemotaxis in immune
CC response. The present sequence is high affinity receptor binding
CC domain of human thrombin. This peptide is used in the exemplification
CC of the invention.

XX

SQ Sequence 23 AA;

Query Match 100.0%; Score 131; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
||||||||||||||||||||||||
Db 1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 5
AAE20159
ID AAE20159 standard; peptide; 23 AA.
XX
AC AAE20159;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human thrombin peptide derivative #2.
XX
KW Cartilage growth; cartilage repair; arthritic joint; traumatic injury;
KW non-proteolytically activated thrombin receptor; NPAR; chondrocyte;
KW therapy; implantation; thrombin peptide; human.
XX
OS Homo sapiens.
XX
PN WO200207748-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US22668.
XX
PR 20-JUL-2000; 2000US-219800P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH, Crowther RS, Stiernberg J, Bergmann J;
XX
DR WPI; 2002-268953/31.
XX
PT Stimulating growth and repair of cartilage, useful for treating e.g.
PT arthritis, by local administration of an agonist of non-proteolytically
PT activated thrombin receptor -
XX
PS Claim 12; Page 25; 28pp; English.
XX
CC The invention relates to a method of stimulating growth and repair of
CC cartilage. The method involves administering to the site, an agonist
CC of non-proteolytically activated thrombin receptor (NPAR). The method
CC is used in human or veterinary medicine for the treatment of arthritic
CC joints and damage/loss of cartilage caused by traumatic injury. Also
CC chondrocytes may be cultured in presence of NPAR agonist to provide
CC cells for implantation at sites requiring growth/repair of cartilage.
CC The present sequence is human thrombin peptide derivative which serves
CC as a NPAR agonist.
XX
SQ Sequence 23 AA;

  Query Match   100.0%; Score 131; DB 23; Length 23;
  Best Local Similarity 100.0%; Pred. No. 3.4e-08;
  Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db   1 AGYKPDEGKRGDACEGDSGGPFV 23

RESULT 6
AAM50858
ID AAM50858 standard; Peptide; 23 AA.
XX
AC AAM50858;
XX
DT 01-MAY-2002 (first entry)
XX
DE Thrombin-derived peptide used to promote cardiac tissue repair.
XX
KW Thrombin; revascularisation; vascular occlusion; tissue repair;
KW vulnerary; vasotropic; cardiant; angiogenesis; restenosis;
KW therapy; human.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Peptide   10..13
FT   /note= "thrombin receptor binding domain"
FT Peptide   12..23
FT   /note= "serine esterase conserved sequence"
XX
PN WO200204008-A2.
XX
PD 17-JAN-2002.
XX
PF 12-JUL-2001; 2001WO-US21944.
XX
PR 12-JUL-2000; 2000US-217583P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Carney DH;
XX
DR WPI; 2002-179665/23.
XX
PT Promoting cardiac tissue repair, stimulating revascularisation,
PT stimulating vascular endothelial cell proliferation, and inhibiting
PT vascular occlusion by using angiogenic thrombin derivative peptide -
XX
PS Claim 4; Page 19; 24pp; English.
XX
CC The present peptide comprises a thrombin-derived peptide, TP508,
CC that includes a thrombin receptor binding domain sequence (see also
CC AAM50856) and a serine esterase conserved sequence (see also
CC AAM50857). The peptide is used in a claimed method for promoting
CC cardiac tissue repair. It is administered during or following
CC cardiac surgery by injection into cardiac tissue, and may be
CC formulated as a sustained release formulation. The thrombin
CC derivative peptide is also used in claimed methods of stimulating
CC revascularisation, stimulating vascular endothelial cell
CC proliferation, inhibiting vascular occlusion, and inhibiting
CC restenosis following balloon angioplasty, in which case it may be
CC coated onto the catheter.
XX
SQ Sequence 23 AA;

```
  Query Match             100.0%;  Score 131;  DB 23;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-08;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db          1 AGYKPDEGKRGDACEGDSGGPFV 23


RESULT 7
AAW99115
ID   AAW99115 standard; protein; 116 AA.
XX
AC   AAW99115;
XX
DT   14-MAY-1999  (first entry)
XX
DE   Human zeta 2 prethrombin 2.
XX
KW   Prothrombin; exosite assay; anticoagulant; blood clot; thrombin;
KW   cardiovascular disease; stroke; haematological disorder.
XX
OS   Homo sapiens.
XX
PN   WO9855130-A1.
XX
PD   10-DEC-1998.
XX
PF   28-MAY-1998;   98WO-US10840.
XX
PR   08-APR-1998;   98US-0081030.
PR   06-JUN-1997;   97US-0048864.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Krishnaswamy S;
XX
DR   WPI; 1999-070237/06.
XX
PT   Exosite assay for agents that inhibit catalytic cleavage of
PT   prothrombin - at sites away from the active site of prothrombinase,
PT   also new inhibitors, potentially useful as anticoagulants
XX
PS   Disclosure; Page 44-45; 61pp; English.
XX
CC   An exosite assay has been developed for inhibition of the catalytic
CC   cleavage of prothrombin (PTh) to thrombin (Th) by prothrombinase (I), at
CC   a site remote from the catalytic site of (I) comprises: (a) preparing a
CC   solution containing 0.05-20 mu M substrate (S), that includes a protease
CC   cleavage site and exosite-binding determinant; 0.05-200 nM factor Va;
CC   30-500 micro M phospholipids (PL); test inhibitor (A) in buffer of pH
CC   7-9, containing 1-10 mM calcium ions but no calcium-chelating agent;
CC   (b) initiating catalytic cleavage of (S) by adding an aliquot of factor
CC   Xa (to final concentration 0.05-200 nM) so that there is an excess of Va
CC   over Xa, forming a S/(I) complex; (c) withdrawing aliquots of the
CC   reaction mixture, quenching them; and (d) assaying for concentration of
CC   Th. Alternatively, in the initial solution S is replaced by the same
CC   concentration of Xa (less than the amount of Va), and reaction is started
CC   by adding S. Also described in the present invention are inhibitors (A')
CC   having IC50 less than 1 mu M identified by this assay. (A') are
CC   potentially useful as a new class of anticoagulants for treatment of
CC   cardiovascular disease, stroke and haematological disorders. The method
CC   is based on the finding that exosite interactions are critical for
CC   substrate specificity in catalytic formation of Th. The present sequence
CC   represents human zeta 2 prethrombin 2.
XX
SQ   Sequence   116 AA;

  Query Match             100.0%;  Score 131;  DB 20;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db         45 AGYKPDEGKRGDACEGDSGGPFV 67


RESULT 8
AAW11545
ID   AAW11545 standard; Protein; 259 AA.
XX
AC   AAW11545;
XX
DT   01-OCT-1997  (first entry)
XX
DE   Human thrombin Asn99 mutant.
XX
KW   Prothrombin; mutant; mutein; platelet aggregation; blood clotting;
KW   coagulation; reduce; decrease; hirudin; heparin; anti-thrombin III;
KW   antagonist; D99N.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..259
FT                   /label= thrombin_Asn99
FT   Misc-difference 99
FT                   /note= "Wild-type Asp residue has been replaced by
FT                         Asn"
XX
PN   WO9641868-A2.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-AT00105.
XX
PR   13-JUN-1995;   95AT-0001006.
XX
PA   (IMMO ) IMMUNO AG.
XX
PI   Eibl J,  Falkner F,  Fischer B,  Mitterer A,  Schlokat U;
XX
```

DR WPI; 1997-065455/06.
XX
PT Prothrombin mutants with reduced clotting activity - useful as
PT antagonists of thrombin inhibitors or for anticoagulant therapy
XX
PS Example 3; Page -; 73pp; German.
XX
CC Prothrombin mutants having one or more changes in amino acid sequence
CC compared with the natural protein and having 0-10% (preferably 0-0.25%)
CC of the activity of the natural protein are claimed, provided that the
CC changes in amino acid sequence do not affect the capacity of the
CC mutants to bind to specific ligands and receptors. The mutants have
CC greatly reduced clotting activity and are useful as antagonists of
CC thrombin inhibitors such as hirudin, heparin and anti-thrombin III.
CC The mutations may also result in changes to the in vivo half-life
CC of prothrombin. The half-life may be reduced to less than 10 minutes
CC or the mutant prothrombin may have an extended half-life of more than
CC 1 hour, making it useful as an anticoagulant and to inhibit side-
CC effects of anti-coagulant treatment. They are converted to inactive
CC thrombin and are able to compete with native, active thrombin for
CC binding to receptors. The present sequence represents the thrombin
CC mutant which is derived by trypsin cleavage of a specifically
CC claimed human prothrombin mutant in which Asp at position 419 is
CC changed to Asn. The thrombin Asn99 mutant was found to have only
CC 0.24% of the activity of wild-type thrombin on a chromogenic
CC substrate.
CC (Note: This sequence does not appear in the specification and has
CC been produced by modifying the wild-type sequence of human
CC prothrombin which appears in figure 1).
XX
SQ Sequence 259 AA;

Query Match 100.0%; Score 131; DB 18; Length 259;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
||||||||||||||||||||||||
Db 188 AGYKPDEGKRGDACEGDSGGPFV 210

RESULT 9
ABP60563
ID ABP60563 standard; protein; 259 AA.
XX
AC ABP60563;
XX
DT 28-MAR-2003 (first entry)
XX
DE Human thrombin variant W215A B-chain.
XX
KW Human; thrombin; W215A; anticoagulant; prothrombin; antithrombotic;
KW thrombus; protein C activation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 229
FT /note= "Wild-type Trp substituted by Ala"
XX
PN WO2002100337-A2.
XX
PD 19-DEC-2002.
XX
PF 07-JUN-2002; 2002WO-US18211.
XX
PR 08-JUN-2001; 2001US-297089P.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Gruber A, Hanson SR, Di Cera E;
XX
DR WPI; 2003-156907/15.
XX
PT New variant thrombin, useful as an antithrombotic agent for inhibiting
PT the formation of a thrombus, for determining the level of protein C
PT activation in a blood sample, or for determining the thrombogenic
PT potential of a patient -
XX
PS Claim 15; Fig 2; 95pp; English.
XX
CC The invention relates to a novel variant human thrombin. The thrombin
CC variant of the invention has anticoagulant activity. The variant thrombin
CC or prothrombin is useful as an antithrombotic agent for inhibiting the
CC formation of a thrombus. The variant thrombin is also useful for
CC determining the level of protein C activation in a blood sample or the
CC thrombogenic potential of a patient. The present sequence represents the
CC B-chain of the thrombin variant W215A.
XX
SQ Sequence 259 AA;

Query Match 100.0%; Score 131; DB 24; Length 259;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGYKPDEGKRGDACEGDSGGPFV 23
||||||||||||||||||||||||
Db 188 AGYKPDEGKRGDACEGDSGGPFV 210

RESULT 10
ABP60565
ID ABP60565 standard; protein; 259 AA.
XX
AC ABP60565;
XX
DT 28-MAR-2003 (first entry)
XX
DE Human thrombin variant W215A/E217A B-chain.
XX
KW Human; thrombin; W215A/E217A; anticoagulant; prothrombin; antithrombotic;
KW thrombus; protein C activation.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 227
FT                   /note= "Wild-type Trp substituted by Ala"
FT   Misc-difference 229
FT                   /note= "Wild-type Glu substituted by Ala"
XX
PN   WO2002100337-A2.
XX
PD   19-DEC-2002.
XX
PF   07-JUN-2002; 2002WO-US18211.
XX
PR   08-JUN-2001; 2001US-297089P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Gruber A,  Hanson SR,  Di Cera E;
XX
DR   WPI; 2003-156907/15.
DR   N-PSDB; ABZ25535.
XX
PT   New variant thrombin, useful as an antithrombotic agent for inhibiting
PT   the formation of a thrombus, for determining the level of protein C
PT   activation in a blood sample, or for determining the thrombogenic
PT   potential of a patient -
XX
PS   Claim 2; Fig 4; 95pp; English.
XX
CC   The invention relates to a novel variant human thrombin. The thrombin
CC   variant of the invention has anticoagulant activity. The variant thrombin
CC   or prothrombin is useful as an antithrombotic agent for inhibiting the
CC   formation of a thrombus. The variant thrombin is also useful for
CC   determining the level of protein C activation in a blood sample or the
CC   thrombogenic potential of a patient. The present sequence represents the
CC   B-chain of the thrombin variant W215A/E217A (WE).
XX
SQ   Sequence   259 AA;

  Query Match           100.0%;  Score 131;  DB 24;  Length 259;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        188 AGYKPDEGKRGDACEGDSGGPFV 210


RESULT 11
AAR74775
ID   AAR74775 standard; Protein; 295 AA.
XX
AC   AAR74775;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Wild-type thrombin.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
DR   N-PSDB; AAQ92455.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Disclosure; Fig 1; 78pp; English.
XX
CC   The sequence reresents wild-type (reference) thrombin.  Mutants
CC   of this sequence (AAR74776-80 and AAR76033-41) have at least 80%
CC   homology with thrombin, and are capable of protein-C activation
CC   without significant fibrinogen clotting activity, and vice versa
CC   (specifically have a ratio of protein-C activity to fibrinogen
CC   clotting activity of less than 0.5 or greater than 2 compared to
CC   thrombin).  The mutant thrombin sequences, produced in recombinant
CC   cell culture or by in vitro methods, and are used to treat
CC   thrombotic conditions, particularly during cardiac bypass surgery
CC   and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;

  Query Match           100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity  100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246


RESULT 12
```

```
AAR74776
ID   AAR74776 standard; Protein; 295 AA.
XX
AC   AAR74776;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin K52A, R233A.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 88
FT                   /note= "Lys in wild-type"
FT   Misc-difference 269
FT                   /note= "Arg in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Claim 22; Page 63/3; 78pp; English.
XX
CC   The mutant thrombin sequence, generated by oligonucleotide-directed
CC   mutagenesis, has at least 80% homology with thrombin, and is
CC   capable of protein-C activation without significant fibrinogen
CC   clotting activity, and vice versa (specifically, it has a ratio
CC   of protein-C activity to fibrinogen clotting activity of less than
CC   0.5 or greater than 2 compared to thrombin).  The mutant thrombin
CC   is produced in recombinant cell culture or by in vitro methods,
CC   and is used to treat thrombotic conditions, particularly during
CC   cardiac bypass surgery and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;
```

```
  Query Match             100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246
```

```
RESULT 13
AAR74777
ID   AAR74777 standard; Protein; 295 AA.
XX
AC   AAR74777;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin E229D.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 265
FT                   /note= "Glu in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Claim 22; Page 63/3; 78pp; English.
XX
CC   The mutant thrombin sequence, generated by oligonucleotide-directed
CC   mutagenesis, has at least 80% homology with thrombin, and is
CC   capable of protein-C activation without significant fibrinogen
CC   clotting activity, and vice versa (specifically, it has a ratio
CC   of protein-C activity to fibrinogen clotting activity of less than
```

```
CC   0.5 or greater than 2 compared to thrombin).  The mutant thrombin
CC   is produced in recombinant cell culture or by in vitro methods,
CC   and is used to treat thrombotic conditions, particularly during
CC   cardiac bypass surgery and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;

  Query Match             100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246


RESULT 14
AAR74778
ID   AAR74778 standard; Protein; 295 AA.
XX
AC   AAR74778;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin E229F.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 265
FT                   /note= "Glu in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Claim 22; Page 63/3; 78pp; English.
XX
CC   The mutant thrombin sequence, generated by oligonucleotide-directed
CC   mutagenesis, has at least 80% homology with thrombin, and is
CC   capable of protein-C activation without significant fibrinogen
CC   clotting activity, and vice versa (specifically, it has a ratio
CC   of protein-C activity to fibrinogen clotting activity of less than
CC   0.5 or greater than 2 compared to thrombin).  The mutant thrombin
CC   is produced in recombinant cell culture or by in vitro methods,
CC   and is used to treat thrombotic conditions, particularly during
CC   cardiac bypass surgery and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;

  Query Match             100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246


RESULT 15
AAR74779
ID   AAR74779 standard; Protein; 295 AA.
XX
AC   AAR74779;
XX
DT   25-MAR-2003  (updated)
DT   04-NOV-1995  (first entry)
XX
DE   Mutant thrombin E229S.
XX
KW   Thrombin; oligonucleotide-directed mutagenesis; procoagulant;
KW   anticoagulant; protein engineering; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 265
FT                   /note= "Glu in wild-type"
FT   Protein         37..295
FT                   /note= "mature protein"
XX
PN   WO9513385-A2.
XX
PD   18-MAY-1995.
XX
PF   14-NOV-1994;   94WO-US13104.
XX
PR   10-JUN-1994;   94US-0258038.
PR   12-NOV-1993;   93US-0152657.
XX
PA   (GILE-) GILEAD SCI.
```

```
XX
PI   Gibbs CS,  Leung LLK,  Tsiang M;
XX
DR   WPI; 1995-194103/25.
XX
PT   Thrombin derivs with segregated pro- and anticoagulant activities -
PT   useful for treating thrombotic disorders but also diagnosis,
PT   treatment of tumours, etc.
XX
PS   Claim 22; Page 63/3; 78pp; English.
XX
CC   The mutant thrombin sequence, generated by oligonucleotide-directed
CC   mutagenesis, has at least 80% homology with thrombin, and is
CC   capable of protein-C activation without significant fibrinogen
CC   clotting activity, and vice versa (specifically, it has a ratio
CC   of protein-C activity to fibrinogen clotting activity of less than
CC   0.5 or greater than 2 compared to thrombin).  The mutant thrombin
CC   is produced in recombinant cell culture or by in vitro methods,
CC   and is used to treat thrombotic conditions, particularly during
CC   cardiac bypass surgery and in cases of septic shock.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   295 AA;

  Query Match             100.0%;  Score 131;  DB 16;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        224 AGYKPDEGKRGDACEGDSGGPFV 246


Search completed: February 11, 2004, 14:53:25
Job time : 49.7097 secs
```

```
                                GenCore version 5.1.6
                           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         February 11, 2004, 14:49:07 ; Search time 15.5806 Seconds
                                             (without alignments)
                                             141.963 Million cell updates/sec

Title:          US-10-050-611-4
Perfect score:  131
Sequence:       1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        PIR_76:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                  SUMMARIES

                          %
Result                  Query
   No.    Score  Match Length DB  ID                      Description
 ---------------------------------------------------------------------------
     1      131  100.0    622  1  TBHU                    thrombin (EC 3.4.2
     2      127   96.9    236  2  C42696                  thrombin (EC 3.4.2
     3      124   94.7    625  1  TBBO                    thrombin (EC 3.4.2
     4      118   90.1    234  2  F42696                  thrombin (EC 3.4.2
     5      113   86.3    235  2  D42696                  thrombin (EC 3.4.2
     6      113   86.3    235  2  E42696                  thrombin (EC 3.4.2
     7      110   84.0    236  2  I42696                  thrombin (EC 3.4.2
     8      109   83.2    239  2  G42696                  thrombin (EC 3.4.2
     9      102   77.9    617  2  S10511                  thrombin (EC 3.4.2
    10      102   77.9    618  2  A35827                  thrombin (EC 3.4.2
    11       89   67.9    235  2  H42696                  thrombin (EC 3.4.2
    12     71.5   54.6    417  1  S00845                  hepsin (EC 3.4.21.
    13       71   54.2    461  1  KXHU                    protein C (activat
```

14	70.5	53.8	482	1	EXRT	coagulation factor
15	70.5	53.8	638	1	KQHUP	plasma kallikrein
16	69.5	53.1	275	2	S40007	trypsin (EC 3.4.21
17	69.5	53.1	1524	2	T30337	polyprotein - Afri
18	68.5	52.3	161	2	I62744	coagulation factor
19	68.5	52.3	488	1	EXHU	coagulation factor
20	68.5	52.3	1019	2	A38738	coagulation factor
21	67.5	51.5	161	2	I48158	coagulation factor
22	67.5	51.5	282	2	I84621	coagulation factor
23	67.5	51.5	459	2	JQ0419	coagulation factor
24	67.5	51.5	475	1	EXCH	coagulation factor
25	67.5	51.5	638	1	KQMSPL	plasma kallikrein
26	67	51.1	225	2	S45356	probable serine pr
27	67	51.1	264	2	S32794	trypsin-like prote
28	66.5	50.8	309	2	B49878	coagulation factor
29	66.5	50.8	1004	2	T30338	oviductin (EC 3.4.
30	65.5	50.0	267	2	S40006	trypsin (EC 3.4.21
31	65.5	50.0	274	2	S35339	trypsin (EC 3.4.21
32	65.5	50.0	275	2	S40005	trypsin (EC 3.4.21
33	65.5	50.0	277	2	S35340	trypsin (EC 3.4.21
34	65.5	50.0	638	1	KQRTPL	plasma kallikrein
35	64.5	49.2	237	2	S55378	serine proteinase
36	64.5	49.2	238	1	TRWV5Y	trypsin-like prote
37	64	48.9	191	2	S54115	complement factor
38	64	48.9	246	1	DBHU	complement factor
39	64	48.9	456	1	KXBO	protein C (activat
40	64	48.9	2616	2	A57096	nudel protein prec
41	63.5	48.5	625	1	KFHU1	coagulation factor
42	63	48.1	461	1	JX0210	protein C (activat
43	62.5	47.7	375	1	A23689	limulus clotting e
44	62.5	47.7	416	1	S33777	hepsin (EC 3.4.21.
45	62.5	47.7	492	1	EXBO	coagulation factor

ALIGNMENTS

RESULT 1
TBHU
thrombin (EC 3.4.21.5) precursor [validated] - human
N;Alternate names: coagulation factor II
N;Contains: prothrombin
C;Species: Homo sapiens (man)
C;Date: 30-Nov-1980 #sequence_revision 22-Jul-1994 #text_change 08-Dec-2000
C;Accession: A29351; A00914; B00914; A37549; A37550; I51952
R;Degen, S.J.F.; Davie, E.W.
Biochemistry 26, 6165-6177, 1987
A;Title: Nucleotide sequence of the gene for human prothrombin.
A;Reference number: A29351; MUID:88077877; PMID:2825773
A;Accession: A29351
A;Molecule type: DNA
A;Residues: 1-622 <DEG>
A;Cross-references: GB:M17262; GB:M33691; NID:g558069; PIDN:AAC63054.1;
PID:g339641
R;Degen, S.J.F.; MacGillivray, R.T.A.; Davie, E.W.
Biochemistry 22, 2087-2097, 1983

A;Title: Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin.
A;Reference number: A00914; MUID:83231469; PMID:6305407
A;Accession: A00914
A;Molecule type: mRNA
A;Residues: 8-163,'N',165-622 <DE2>
A;Cross-references: GB:V00595; GB:J00307; NID:g37128; PIDN:CAA23842.1;
PID:g1335344
A;Accession: B00914
A;Molecule type: DNA
A;Residues: 188-311 <DE3>
R;Walz, D.A.; Hewett-Emmett, D.; Seegers, W.H.
Proc. Natl. Acad. Sci. U.S.A. 74, 1969-1972, 1977
A;Reference number: A37549; MUID:77193964; PMID:266717
A;Accession: A37549
A;Molecule type: protein
A;Residues: 44-118,'N',120,'S',122-163,'I',165-175,'A',177-182,'T',184-193,'MV',196-308,'EE',309-314 <WAL>
R;Butkowski, R.J.; Elion, J.; Downing, M.R.; Mann, K.G.
J. Biol. Chem. 252, 4942-4957, 1977
A;Title: Primary structure of human prethrombin 2 and alpha-thrombin.
A;Reference number: A37550; MUID:77207112; PMID:873923
A;Accession: A37550
A;Molecule type: protein
A;Residues: 315-334,'N',336-348,'N',350-368,'N',370-397,'N',399-413,'N',415-484,'N',486-493,'G',495-503,'Y',505-508,'S',510,'V',512-513,'D',515-528,'AL',531,'Q',533-622 <BUT>
R;Rabiet, M.J.; Blashill, A.; Furie, B.; Furie, B.C.
J. Biol. Chem. 261, 13210-13215, 1986
A;Reference number: A37551; MUID:87008532; PMID:3759958
A;Contents: annotation; activation cleavages
R;MacGillivray, R.T.; Irwin, D.M.; Guinto, E.R.; Stone, J.C.
Ann. N. Y. Acad. Sci. 485, 73-79, 1986
A;Title: Recombinant genetic approaches to functional mapping of thrombin.
A;Reference number: I51952; MUID:87182874; PMID:3471151
A;Accession: I51952
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2,'RI',5-100 <RES>
A;Cross-references: GB:M33031; NID:g190723; PIDN:AAA60220.1; PID:g190724
C;Comment: Thrombin, which cleaves bonds after Arg and Lys, converts fibrinogen to fibrin and activates factors V, VIII, XIII, and, in complex with thrombomodulin, protein C.
C;Comment: Prothrombin is activated on the surface of a phospholipid membrane that binds the amino end of prothrombin and factors Va and Xa in calcium-dependent interactions. The activation peptide(s) can be removed either by factor Xa or thrombin; the cleavage into light and heavy chains is by factor Xa. It is not known whether one or two smaller activation peptides, with additional cleavage after 314-Arg, are released in natural blood clotting.
C;Comment: The cleavage after Arg-198, observed in vitro, does not occur in plasma.
C;Comment: The gamma-carboxyglutamyl residues bind calcium ions, result from the carboxylation of glutamyl residues by microsomal vitamin K-dependent carboxylase, and are necessary for calcium-dependent interaction with the negatively charged phospholipid membrane surface.
C;Comment: The prothrombin precursor is synthesized in the liver.
C;Genetics:

A;Gene: GDB:F2
A;Cross-references: GDB:119894; OMIM:176930
A;Map position: 11p11-11q12
A;Introns: 27/1; 80/3; 89/1; 106/1; 141/2; 187/1; 292/1; 335/1; 377/2; 433/2; 491/2; 552/1; 575/3
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: acute phase; blood coagulation; calcium binding; carboxyglutamic acid; duplication; glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-43/Domain: propeptide #status predicted <PRO>
F;28-87/Domain: Gla domain homology <GLA>
F;44-622/Product: prothrombin #status experimental <MAT>
F;44-327/Domain: activation peptide #status experimental <APT>
F;108-186/Domain: kringle homology <KR1>
F;213-291/Domain: kringle homology <KR2>
F;328-363/Product: thrombin light chain #status experimental <LCH>
F;364-622/Product: thrombin heavy chain #status experimental <HCH>
F;364-613/Domain: trypsin homology <TRY>
F;49,50,57,59,62,63,68,69,72,75/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental
F;60-65,90-103,108-186,129-169,157-181,213-291,234-274,262-286/Disulfide bonds: #status predicted
F;121,143/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;336-482,536-550,564-594/Disulfide bonds: #status predicted
F;391-407/Disulfide bonds: #status experimental
F;406,462/Active site: His, Asp #status predicted
F;416/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;568/Active site: Ser #status experimental

  Query Match 100.0%; Score 131; DB 1; Length 622;
  Best Local Similarity 100.0%; Pred. No. 1.9e-10;
  Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        551 AGYKPDEGKRGDACEGDSGGPFV 573
```

RESULT 2
C42696
thrombin (EC 3.4.21.5) B chain - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 17-Mar-1999
C;Accession: C42696
R;Banfield, D.K.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 2779-2783, 1992
A;Title: Partial characterization of vertebrate prothrombin cDNAs: amplification and sequence analysis of the B chain of thrombin from nine different species.
A;Reference number: A42696; MUID:92212913; PMID:1557383
A;Accession: C42696
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-236 <BAN>
A;Cross-references: GB:M81396
C;Superfamily: thrombin; Gla domain homology; kringle homology; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-227/Domain: trypsin homology (fragment) <TRY>

  Query Match 96.9%; Score 127; DB 2; Length 236;
  Best Local Similarity 95.7%; Pred. No. 2.6e-10;
  Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||:||||||||||||||||
Db        165 AGYKPEEGKRGDACEGDSGGPFV 187
```

Search completed: February 11, 2004, 14:56:57
Job time : 15.5806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:36:52 ; Search time 9.64516 Seconds
(without alignments)
112.141 Million cell updates/sec

Title: US-10-050-611-4
Perfect score: 131
Sequence: 1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	131	100.0	622	1	THRB_HUMAN	P00734 homo sapien
2	124	94.7	625	1	THRB_BOVIN	P00735 bos taurus
3	102	77.9	617	1	THRB_RAT	P18292 rattus norv
4	102	77.9	618	1	THRB_MOUSE	P19221 mus musculu
5	73.5	56.1	290	1	MPN_HUMAN	Q9bqr3 homo sapien
6	71.5	54.6	417	1	HEPS_HUMAN	P05981 homo sapien
7	71.5	54.6	436	1	HEPS_MOUSE	O35453 mus musculu
8	71	54.2	161	1	PRTC_MACMU	Q28506 macaca mula
9	71	54.2	461	1	PRTC_HUMAN	P04070 homo sapien
10	70.5	53.8	638	1	KAL_HUMAN	P03952 homo sapien
11	70	53.4	281	1	TRYZ_DROER	P54630 drosophila
12	69.5	53.1	275	1	TRY3_ANOGA	P35037 anopheles g
13	68.5	52.3	488	1	FA10_HUMAN	P00742 homo sapien
14	68.5	52.3	1019	1	LFC_CARRO	Q26422 carcinoscor
15	68.5	52.3	1019	1	LFC_TACTR	P28175 tachypleus
16	68	51.9	458	1	PRTC_RABIT	Q28661 oryctolagus
17	67.5	51.5	282	1	FA9_RAT	P16296 rattus norv
18	67.5	51.5	459	1	FA9_MOUSE	P16294 mus musculu
19	67.5	51.5	475	1	FA10_CHICK	P25155 gallus gall
20	67.5	51.5	638	1	KAL_MOUSE	P26262 mus musculu
21	67	51.1	256	1	KLKF_HUMAN	Q9h2r5 homo sapien
22	67	51.1	264	1	VDP_BOMMO	Q07943 bombyx mori
23	66.5	50.8	455	1	TMS5_MOUSE	Q9er04 mus musculu
24	66.5	50.8	457	1	TMS5_HUMAN	Q9h3s3 homo sapien
25	65.5	50.0	267	1	TRY7_ANOGA	P35041 anopheles g
26	65.5	50.0	274	1	TRY1_ANOGA	P35035 anopheles g
27	65.5	50.0	275	1	TRY4_ANOGA	P35038 anopheles g
28	65.5	50.0	277	1	TRY2_ANOGA	P35036 anopheles g
29	65.5	50.0	638	1	KAL_RAT	P14272 rattus norv
30	65	49.6	157	1	PRTC_CANFA	Q28278 canis famil
31	65	49.6	157	1	PRTC_CAPHI	Q28315 capra hircu
32	65	49.6	157	1	PRTC_FELCA	Q28412 felis silve
33	65	49.6	157	1	PRTC_HORSE	Q28380 equus cabal
34	65	49.6	459	1	PRTC_PIG	Q9glp2 sus scrofa
35	64.5	49.2	238	1	TRY5_AEDAE	P29787 aedes aegyp
36	64.5	49.2	422	1	DES1_HUMAN	Q9ul52 homo sapien
37	64.5	49.2	490	1	FA10_RABIT	O19045 oryctolagus
38	64	48.9	253	1	CFAD_HUMAN	P00746 homo sapien
39	64	48.9	259	1	CFAD_PIG	P51779 sus scrofa
40	64	48.9	456	1	PRTC_BOVIN	P00745 bos taurus
41	64	48.9	875	1	NETR_HUMAN	P56730 homo sapien
42	64	48.9	2616	1	NDL_DROME	P98159 drosophila
43	63.5	48.5	625	1	FA11_HUMAN	P03951 homo sapien
44	63	48.1	256	1	TRYE_DROER	P54627 drosophila
45	63	48.1	461	1	PRTC_MOUSE	P33587 mus musculu

ALIGNMENTS

RESULT 1
THRB_HUMAN
ID THRB_HUMAN STANDARD; PRT; 622 AA.
AC P00734;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Prothrombin precursor (EC 3.4.21.5) (Coagulation factor II).
GN F2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88077877; PubMed=2825773;
RA Degen S.J.F., Davie E.W.;
RT "Nucleotide sequence of the gene for human prothrombin.";
RL Biochemistry 26:6165-6177(1987).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT MET-165.
RA Rieder M.J., Armel T.Z., Carrington D.P., Chung M.-W., Lee K.L.,
RA Ozuna M., Poel C.L., Toth E.J., Yi Q., Nickerson D.A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.

RN [3]
RP SEQUENCE OF 8-622 FROM N.A.
RX MEDLINE=83231469; PubMed=6305407;
RA Degen S.J.F., McGillivray R.T.A., Davie E.W.;
RT "Characterization of the complementary deoxyribonucleic acid and gene
RT coding for human prothrombin.";
RL Biochemistry 22:2087-2097(1983).
RN [4]
RP SEQUENCE OF 44-314.
RX MEDLINE=77193964; PubMed=266717;
RA Walz D.A., Hewett-Emmett D., Seegers W.H.;
RT "Amino acid sequence of human prothrombin fragments 1 and 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 74:1969-1972(1977).
RN [5]
RP SEQUENCE OF 315-622.
RX MEDLINE=77207112; PubMed=873923;
RA Butkowski R.J., Elion J., Downing M.R., Mann K.G.;
RT "Primary structure of human prethrombin 2 and alpha-thrombin.";
RL J. Biol. Chem. 252:4942-4957(1977).
RN [6]
RP PROCESSING.
RX MEDLINE=87008532; PubMed=3759958;
RA Rabiet M.J., Blashill A., Furie B., Furie B.C.;
RT "Prothrombin fragment 1 X 2 X 3, a major product of prothrombin
RT activation in human plasma.";
RL J. Biol. Chem. 261:13210-13215(1986).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=90059942; PubMed=2583108;
RA Bode W., Mayr I., Baumann U., Huber R., Stone S.R., Hofsteenge J.;
RT "The refined 1.9 A crystal structure of human alpha-thrombin:
RT interaction with D-Phe-Pro-Arg chloromethylketone and significance of
RT the Tyr-Pro-Pro-Trp insertion segment.";
RL EMBO J. 8:3467-3475(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=90327074; PubMed=2374926;
RA Rydel T.J., Ravichandran K.G., Tulinsky A., Bode W., Huber R.,
RA Roitsch C., Fenton J.W. II;
RT "The structure of a complex of recombinant hirudin and human alpha-
RT thrombin.";
RL Science 249:277-280(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=94350942; PubMed=8071320;
RA Rydel T.J., Yin M., Padmanabhan K.P., Blankenship D.T., Cardin A.D.,
RA Correa P.E., Fenton J.W. II, Tulinsky A.;
RT "Crystallographic structure of human gamma-thrombin.";
RL J. Biol. Chem. 269:22000-22006(1994).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=97357286; PubMed=9214615;
RA van de Locht A., Bode W., Huber R., le Bonniec B.F., Stone S.R.,
RA Esmon C.T., Stubbs M.T.;
RT "The thrombin E192Q-BPTI complex reveals gross structural
RT rearrangements: implications for the interaction with antithrombin
RT and thrombomodulin.";
RL EMBO J. 16:2977-2984(1997).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 328-601.
RX MEDLINE=99162521; PubMed=10051558;
RA Guinto E.R., Caccia S., Rose T., Fuetterer K., Waksman G., di Cera E.;
RT "Unexpected crucial role of residue 225 in serine proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1852-1857(1999).
RN [12]
RP VARIANT BARCELONA.
RX MEDLINE=87033739; PubMed=3771562;
RA Rabiet M.-J., Furie B.C., Furie B.;
RT "Molecular defect of prothrombin Barcelona. Substitution of cysteine
RT for arginine at residue 273.";
RL J. Biol. Chem. 261:15045-15048(1986).
RN [13]
RP VARIANT FRANKFURT.
RX MEDLINE=95313001; PubMed=7792730;
RA Degen S.J.F., McDowell S.A., Sparks L.M., Scharrer I.;
RT "Prothrombin Frankfurt: a dysfunctional prothrombin characterized by
RT substitution of Glu-466 by Ala.";
RL Thromb. Haemost. 73:203-209(1995).
RN [14]
RP VARIANTS HIMI-1 AND HIMI-2.
RX MEDLINE=93043342; PubMed=1421398;
RA Morishita E., Saito M., Kumabashiri I., Asakura H., Matsuda T.,
RA Yamaguchi K.;
RT "Prothrombin Himi: a compound heterozygote for two dysfunctional
RT prothrombin molecules (Met-337-->Thr and Arg-388-->His).";
RL Blood 80:2275-2280(1992).
RN [15]
RP VARIANT PADUA-1.
RX MEDLINE=95169898; PubMed=7865694;
RA James H.L., Kim D.J., Zheng D.-Q., Girolami A.;
RT "Prothrombin Padua I: incomplete activation due to an amino acid
RT substitution at a factor Xa cleavage site.";
RL Blood Coagul. Fibrinolysis 5:841-844(1994).
RN [16]
RP VARIANT QUICK-1.
RX MEDLINE=89207504; PubMed=3242619;
RA Henriksen R.A., Mann K.G.;
RT "Identification of the primary structural defect in the dysthrombin
RT thrombin Quick I: substitution of cysteine for arginine-382.";
RL Biochemistry 27:9160-9165(1988).
RN [17]
RP VARIANT QUICK-2.
RX MEDLINE=89247398; PubMed=2719946;
RA Henriksen R.A., Mann K.G.;
RT "Substitution of valine for glycine-558 in the congenital dysthrombin
RT thrombin Quick II alters primary substrate specificity.";
RL Biochemistry 28:2078-2082(1989).
RN [18]
RP VARIANT SALAKTA.
RX MEDLINE=92378975; PubMed=1354985;
RA Miyata T., Aruga R., Umeyama H., Bezeaud A., Guillin M.-C.,
RA Iwanaga S.;
RT "Prothrombin Salakta: substitution of glutamic acid-466 by alanine
RT reduces the fibrinogen clotting activity and the esterase activity.";

RL Biochemistry 31:7457-7462(1992).
RN [19]
RP VARIANT TOKUSHIMA.
RX MEDLINE=87185407; PubMed=3567158;
RA Miyata T., Morita T., Inomoto T., Kawauchi S., Shirakami A.,
RA Iwanaga S.;
RT "Prothrombin Tokushima, a replacement of arginine-418 by tryptophan
RT that impairs the fibrinogen clotting activity of derived thrombin
RT Tokushima.";
RL Biochemistry 26:1117-1122(1987).
RN [20]
RP VARIANT TOKUSHIMA.
RX MEDLINE=87101511; PubMed=3801671;
RA Inomoto T., Shirakami A., Kawauchi S., Shigekiyo T., Saito S.,
RA Miyoshi K., Morita T., Iwanaga S.;
RT "Prothrombin Tokushima: characterization of dysfunctional thrombin
RT derived from a variant of human prothrombin.";
RL Blood 69:565-569(1987).
RN [21]
RP VARIANT TOKUSHIMA.
RX MEDLINE=92256895; PubMed=1349838;
RA Iwahana H., Yoshimoto K., Shigekiyo T., Shirakami A., Saito S.,
RA Itakura M.;
RT "Detection of a single base substitution of the gene for prothrombin
RT Tokushima. The application of PCR-SSCP for the genetic and molecular
RT analysis of dysprothrombinemia.";
RL Int. J. Hematol. 55:93-100(1992).
RN [22]
RP VARIANT TYPE-3.
RX MEDLINE=83204687; PubMed=6405779;
RA Board P.G., Shaw D.C.;
RT "Determination of the amino acid substitution in human prothrombin
RT type 3 (157 Glu leads to Lys) and the localization of a third
RT thrombin cleavage site.";
RL Br. J. Haematol. 54:245-254(1983).
RN [23]
RP VARIANTS MET-165 AND THR-386.
RX MEDLINE=99318093; PubMed=10391209;
RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA Lander E.S.;
RT "Characterization of single-nucleotide polymorphisms in coding regions
RT of human genes.";
RL Nat. Genet. 22:231-238(1999).
RN [24]
RP ERRATUM.
RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA Lander E.S.;
RL Nat. Genet. 23:373-373(1999).
CC -!- FUNCTION: THROMBIN, WHICH CLEAVES BONDS AFTER ARG & LYS, CONVERTS
CC FIBRINOGEN TO FIBRIN AND ACTIVATES FACTORS V, VII, VIII, XIII,
CC AND, IN COMPLEX WITH THROMBOMODULIN, PROTEIN C.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Gly; activates
CC fibrinogen to fibrin and releases fibrinopeptide A and B.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC -!- PTM: THE GAMMA-CARBOXYGLUTAMYL RESIDUES, WHICH BIND CALCIUM IONS,
CC RESULT FROM THE CARBOXYLATION OF GLUTAMYL RESIDUES BY A MICROSOMAL
CC ENZYME, THE VITAMIN K-DEPENDENT CARBOXYLASE. THE MODIFIED RESIDUES
CC ARE NECESSARY FOR THE CA-DEPENDENT INTERACTION WITH A NEGATIVELY
CC CHARGED PHOSPHOLIPID SURFACE,WHICH IS ESSENTIAL FOR THE CONVERSION

Query Match 100.0%; Score 131; DB 1; Length 622;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGYKPDEGKRGDACEGDSGGPFV 23
              |||||||||||||||||||||||
Db        551 AGYKPDEGKRGDACEGDSGGPFV 573
```

Search completed: February 11, 2004, 14:54:04
Job time : 9.64516 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 14:47:57 ; Search time 39.3226 Seconds
(without alignments)
150.936 Million cell updates/sec

Title: US-10-050-611-4
Perfect score: 131
Sequence: 1 AGYKPDEGKRGDACEGDSGGPFV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	127	96.9	235	6	Q28731	Q28731 oryctolagus
2	118	90.1	235	13	Q90387	Q90387 cynops pyrr
3	113	86.3	235	13	Q91004	Q91004 gecko gecko
4	113	86.3	607	13	Q91001	Q91001 gallus gall
5	113	86.3	608	13	Q9PTW7	Q9ptw7 struthio ca
6	109	83.2	239	13	Q91218	Q91218 oncorhynchu
7	105	80.2	420	13	Q90504	Q90504 eptatretus
8	98	74.8	172	13	Q9DFD1	Q9dfd1 oncorhynchu
9	92	70.2	234	13	Q90244	Q90244 acipenser t
10	72.5	55.3	389	13	Q9PVX7	Q9pvx7 xenopus lae
11	72.5	55.3	974	13	Q90WD8	Q90wd8 bufo japoni
12	71.5	54.6	435	11	Q9CW97	Q9cw97 mus musculu
13	71.5	54.6	799	11	Q9DBI0	Q9dbi0 mus musculu
14	71.5	54.6	802	4	Q8IUE2	Q8iue2 homo sapien
15	71.5	54.6	811	4	Q8IU80	Q8iu80 homo sapien
16	71	54.2	195	4	Q8J008	Q8j008 homo sapien
17	71	54.2	195	4	Q8J007	Q8j007 homo sapien
18	71	54.2	195	4	Q8J006	Q8j006 homo sapien
19	71	54.2	195	4	Q8IXB4	Q8ixb4 homo sapien
20	71	54.2	211	4	Q8J009	Q8j009 homo sapien
21	70.5	53.8	161	11	Q63109	Q63109 rattus norv
22	70.5	53.8	259	5	Q9XY61	Q9xy61 ctenocephal
23	70.5	53.8	267	5	Q9BK47	Q9bk47 luidia foli
24	70.5	53.8	481	11	O54740	O54740 mus musculu
25	70.5	53.8	481	11	Q99L32	Q99132 mus musculu
26	70.5	53.8	481	11	O88947	O88947 mus musculu
27	70.5	53.8	482	11	Q63207	Q63207 rattus norv
28	70	53.4	378	5	Q8SY50	Q8sy50 drosophila
29	69.5	53.1	200	11	Q924U6	Q924u6 mus musculu
30	69.5	53.1	1524	13	Q91674	Q91674 xenopus lae
31	68.5	52.3	161	6	Q28511	Q28511 macaca mula
32	68.5	52.3	236	5	Q9TYH3	Q9tyh3 schistosoma
33	68.5	52.3	488	5	Q9TYH4	Q9tyh4 schistosoma
34	68.5	52.3	766	4	Q8NBY4	Q8nby4 homo sapien
35	68.5	52.3	1019	5	Q8T9S1	Q8t9s1 tachypleus
36	68.5	52.3	1083	5	Q26423	Q26423 carcinoscor
37	68	51.9	686	13	Q9DGC2	Q9dgc2 cyprinus ca
38	67.5	51.5	156	5	O16007	O16007 schistosoma
39	67.5	51.5	161	11	Q60546	Q60546 mesocricetu
40	67.5	51.5	264	5	O02569	O02569 culex quinq
41	67.5	51.5	328	11	Q8BJR6	Q8bjr6 mus musculu
42	67.5	51.5	370	5	Q9VA44	Q9va44 drosophila
43	67.5	51.5	387	5	Q9XY57	Q9xy57 ctenocephal
44	67.5	51.5	474	13	Q8JHC8	Q8jhc8 brachydanio
45	67.5	51.5	638	11	Q8R0P5	Q8r0p5 mus musculu

Search completed: February 11, 2004, 14:56:05
Job time : 39.3226 secs